

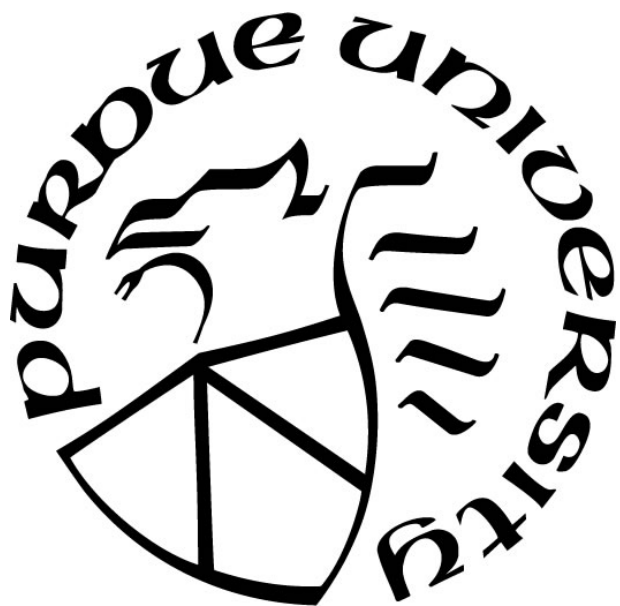
MEASURING AND MODELING OF PHENYLPROPANOID METABOLIC FLUX IN ARABIDOPSIS

by
Peng Wang

A Dissertation

*Submitted to the Faculty of Purdue University
In Partial Fulfillment of the Requirements for the degree of*

Doctor of Philosophy



Department of Biochemistry
West Lafayette, Indiana
May 2019

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Clint Chapple, Chair

Department of Biochemistry

Dr. Natalia Dudareva

Department of Biochemistry

Dr. John Morgan

Davison School of Chemical Engineering

Dr. Brian Dilkes

Department of Biochemistry

Approved by:

Dr. Andrew Mesecar

Head of the Graduate Program

ACKNOWLEDGEMENTS

On my way of education, I always feel blessed that I, born in a small town, grow into an independent scientist pursuing the cutting-edge research. In this journey, I thank my advisor Dr. Clint Chapple the most. It is his invaluable mentorship and encouragement that have made me achieve the accomplishments today. Through the years, he has guided me with his profound wisdom and unique vision as a scientist and with his bright optimism and endless support as a mentor. The positive way he interprets results has stimulated me to view things from different angles. The high expectation he holds on students has pushed me to exert my best endeavor. All I have learnt from him will benefit me for my whole life.

I would like to appreciate Dr. John Morgan, Dr. Natalia Dudareva, Dr. Brian Dilkes, and Dr. David Rhodes, who served on my committee. On the monthly group meetings for years Dr. Natalia Dudareva and Dr. John Morgan have tutored me with inspiration and criticism. Dr. Brian Dilkes' advises on bioinformatics have been very helpful for my transcriptome research work.

I cannot be more grateful to my close collaborators, Longyun Guo, Rohit Jaini, and Shaunak Ray, with whom together I have conducted the interdisciplinary projects. They have taught me not only how to think about scientific problems with an engineer's mind, but also how to work efficiently on a team. I greatly appreciate their patience and collaboration, which have made my PhD special and memorable.

I own many thanks to all the Chapple lab members, past and present, for their help and accompany. Thanks to the industrious efforts Jo Cusumano devoted, the lab is an incredibly organized environment to work even after she has retired. In my junior years, I was encouraged a lot by Nicholas Bonawitz, Yi Li, Nickolas Anderson, Jeong Im Kim, and Whitney Dolan. They patiently listened to my broken English during journal club, critically pointed out my problems at lab meetings, generously congratulated me on my success. They are not just peers and friends, but also role models for me to look up to.

I would like to thank my families and friends for their never-ending love and constructive support. My parents valued my education since I was a kid and promoted me to go further and further to chase my dreams. This thesis is dedicated to them for their care and encouragement.

TABLE OF CONTENTS

LIST OF TABLES.....	7
LIST OF FIGURES	8
STATEMENT OF PUBLISHED AND COLLABORATIVE WORK	10
ABSTRACT.....	11
CHAPTER 1. GENETIC MANIPULATION OF LIGNOCELLULOSIC BIOMASS FOR BIOENERGY.....	13
1.1 Abstract.....	13
1.2 Introduction	13
1.3 Identification of new genes for lignin manipulation	14
1.4 Modifications of lignin monomers	15
1.5 Regulations of lignocellulose biosynthesis	17
1.6 Engineering of energy crops for lignocellulosic utilization	18
1.7 Conclusions	19
1.8 References	19
CHAPTER 2. A ¹³ C ISOTOPE LABELING METHOD FOR THE MEASUREMENT OF LIGNIN METABOLIC FLUX IN ARABIDOPSIS STEMS.....	31
2.1 Abstract.....	31
2.1.1 Background	31
2.1.2 Results	31
2.1.3 Conclusions	31
2.1.4 Keywords	31
2.2 Background.....	32
2.3 Results	34
2.3.1 Establishment of a stem feeding system using wild-type Arabidopsis	34
2.3.2 Phenylpropanoid metabolism is not homogeneous along the stem.....	35
2.3.3 Soluble phenylpropanoids are rapidly labeled in stems fed with [¹³ C ₆]-Phe	36
2.3.4 Isotopic label is incorporated into lignin in stems fed with [¹³ C ₆]-Phe.....	38
2.4 Discussion.....	39
2.5 Conclusions	42

2.6	Methods	43
2.6.1	Plant material and growth conditions.....	43
2.6.2	Feeding	43
2.6.3	Stem growth measurement	43
2.6.4	LC-MS/MS analysis of soluble metabolites.....	44
2.6.5	DFRC/GC/FID and DFRC/GC/MS	44
2.6.6	Acetyl bromide lignin analysis.....	45
2.6.7	PAL and 4CL enzyme assays.....	45
2.7	References	46
CHAPTER 3. DYNAMIC MODELING OF SUBCELLULAR PHENYLPROPANOID METABOLISM IN ARABIDOPSIS LIGNIFYING CELLS		66
3.1	Abstract.....	66
3.1.1	Keywords	66
3.2	Introduction	66
3.3	Materials and Methods	68
3.3.1	Feeding study in wild type Arabidopsis stem	68
3.3.2	LC-MS/MS analysis	69
3.3.3	Lignin analysis	69
3.3.4	Enzyme assays.....	70
3.3.5	Kinetic modeling	71
3.3.6	Parameter identification	73
3.3.7	Model comparison with information criteria	74
3.3.8	Metabolic control analysis	74
3.4	Results	74
3.4.1	Metabolic profiling of wild-type Arabidopsis basal stem fed with ring labeled [¹³ C ₆]- Phe.....	74
3.4.2	Base kinetic model construction and parameterization.....	75
3.4.3	Incorporation of vacuolar storage of Phe and <i>p</i> -coumarate significantly improve model performance.....	76
3.4.4	No <i>in silico</i> evidence for existence of other metabolite-enzyme interactions.....	77
3.4.5	Refined model unravels subcellular pathway dynamics during feeding studies.....	78

3.4.6	Metabolic control analysis determines the relative enzymatic control of flux in the general phenylpropanoid pathway	79
3.5	Discussion.....	80
3.6	References	82
CHAPTER 4. THE TRANSCRIPTS AND METABOLITES NETWORK IN RESPONSE TO LIGNIN BIOSYNTHESIS PERTURBATIONS IN ARABIDOPSIS		101
4.1	Introduction	101
4.2	Results	103
4.2.1	Global transcriptome analysis in Arabidopsis with perturbed lignification.....	103
4.2.2	Accumulation of phenylpropanoids changed in mutants	105
4.2.3	Transcription of genes involved in lignification altered in mutants.....	106
4.2.4	Analysis of differentially expressed genes (DEGs) in phenylpropanoid mutants identified multiple mis-regulated processes	108
4.2.5	Comparison of metabolites and DEGs in subgroups of mutants	109
4.3	Discussions	113
4.4	Methods	118
4.4.1	Plant materials and growth.....	118
4.4.2	RNA-seq.....	119
4.4.3	Statistical analysis of RNA-seq data	119
4.4.4	Soluble metabolite analysis on LC/MS-MS.....	120
4.5	References	120

LIST OF TABLES

Table 3.1 Subcellular metabolite profiles under different feeding conditions predicted by the refined mode	88
Table 3.2 Final parameter list estimated by the refined model.....	89
Table 3.3 AIC scores of the models relative to the one without any potential metabolite-enzyme interactions.	90
Table 3.4 Flux control coefficients of pathway enzymes	91
Table 4.1 The rank of expression of lignin biosynthetic genes in wild type samples	127
Table 4.2 Concentrations of 18 phenylpropanoids from wild type and mutants.	128
Table 4.3 Gene ontology enrichment of gene clusters from the hierarchical clustering results.	128
Table 4.4 Common DEGs and enriched gene ontology from subgroups of mutants.	128
Table 4.5 Primer sequences for genotyping the Arabidopsis mutants.....	129

LIST OF FIGURES

Figure 1.1 A simplified pathway illustrating the enzymes and metabolites involved in lignin biosynthesis.....	27
Figure 1.2 A simplified model illustrating that manipulations of lignin biosynthetic genes change pathway flux.....	28
Figure 1.3 A model of lignin polymer illustrates the native and modified monomers and structural units.....	29
Figure 1.4 Knocking-out <i>MED5a</i> and <i>MED5b</i> substantially rescues the arrested growth of <i>Arabidopsis ref8-2</i> mutant	30
Figure 2.1 A simplified pathway illustrating lignin biosynthesis.....	53
Figure 2.2 Excised stems incubated in tubes with MS medium.....	54
Figure 2.3 Growth of excised <i>Arabidopsis</i> stems in liquid medium.....	55
Figure 2.4 Medium absorbed by the excised stems during the feeding process.....	56
Figure 2.5 Labeling incorporation into lignin in stems fed with [¹³ C ₆]-Phe.....	57
Figure 2.6 Phenylpropanoid metabolism along wild-type <i>Arabidopsis</i> stems.....	58
Figure 2.7 PAL and 4CL activities in <i>Arabidopsis</i> stems over a day-night cycle.....	59
Figure 2.8 Labeling percentage of soluble phenylpropanoids over the feeding time course.....	60
Figure 2.9 Hierarchical clustering of labeling percentage profiles of soluble phenylpropanoids.....	61
Figure 2.10 Profiles of labeled and unlabeled phenylpropanoids over the feeding time course.....	62
Figure 2.11 Profiles of total phenylpropanoids over the feeding time course.....	63
Figure 2.12 PAL and 4CL activities in the stems over the feeding course.....	64
Figure 2.13 Analysis of lignin monomers in excised <i>Arabidopsis</i> stems fed with [¹³ C ₆]-Phe.....	65
Figure 3.1 Lignin biosynthesis in <i>Arabidopsis</i>	92
Figure 3.2 Data fitting comparison between base and refined models.....	93
Figure 3.3 Total [¹³ C ₆]-lignin in stems fed with [¹³ C ₆]-Phe.....	94
Figure 3.4 Metabolite-enzyme interactions and vacuole storage mechanisms considered in the general phenylpropanoid pathway.....	95
Figure 3.5 Reduction in the AIC score relative to the base model without vacuole storage.....	96
Figure 3.6 Measured and estimated PAL and 4CL activities in wild type stems over the feeding period.....	97

Figure 3.7 AIC scores of the models relative to one without cinnamate and <i>p</i> -coumaroyl CoA mass balances.	98
Figure 3.8 Pathway flux distribution at 240 min for different feeding treatments.	99
Figure 3.9 Flux distribution of the phenylpropanoid pathway during 3 mM ¹³ C ₆ Phe feeding condition.	100
Figure 4.1 Growth phenotype of 28-day-old wild-type Arabidopsis and plants with perturbed lignification.	130
Figure 4.2 Association of the 50 genes with the highest expression.	131
Figure 4.3 PCA analysis of gene expression profiles in wild type and the perturbed plants.....	132
Figure 4.4 PCA analysis of expressed genes in RNA-seq data from individual samples of wild type, <i>ref3-2</i> , and <i>ref3-3</i>	133
Figure 4.5 PCA analysis of expressed genes in RNA-seq data from individual samples of wild type, <i>med5a/b</i> , <i>med5a/b ref8</i> , and <i>med5ab ref2</i>	134
Figure 4.6 Comparison of gene expression profiles among all the samples.....	135
Figure 4.7 Metabolic profiles from wild type and the perturbed plants.	136
Figure 4.8 Expression of lignin biosynthetic genes in the RNA-seq data.	137
Figure 4.9 Expression of genes involved in shikimate biosynthesis in the RNA-seq data.....	138
Figure 4.10 Expression of genes involved in regulating phenylpropanoid metabolism.....	139
Figure 4.11 Numbers of substantial DEGs in each perturbed plant compared to wild type.....	140
Figure 4.12 Pearson's correlation (r) of gene expression profiles.	141
Figure 4.13 Hierarchical clustering analysis of substantial DEGs from 13 mutants compared with wild type.....	142
Figure 4.14 Comparison of metabolites and transcripts in wild type, <i>pal1 pal2</i> , <i>fah1</i> , and <i>ref2</i>	143
Figure 4.15 Comparison of metabolites and transcripts in wild type, <i>ref3-2</i> , and <i>ref3-3</i>	145
Figure 4.16 Comparison of metabolites and transcripts in wild type, <i>4cl1</i> , and <i>cse-2</i>	146
Figure 4.17 Comparison of metabolites and transcripts in wild type, <i>cadC cadD</i> , and <i>C4H:F5H</i>	147
Figure 4.18 Comparison of transcripts and metabolites in wild type, <i>med5a/b</i> , <i>med5a/b ref8</i> , and <i>med5a/b ref2</i>	148
Figure 4.19 Comparison of metabolites and transcripts in wild type, <i>fah1</i> , <i>C4H:F5H</i> , and <i>ref8 fah1 SmF5H</i>	150

STATEMENT OF PUBLISHED AND COLLABORATIVE WORK

Chapter 1 has been published as a review in *Current Opinion in Chemical Biology*, Volume 29, December 2015, under doi 10.1016/j.cbpa.2015.08.006.

Chapter 2 has been published in *Plant Methods*, Volume 14, December 2018, under doi 10.1186/s13007-018-0318-3. I wrote the manuscript and Longyun Guo contributed equally for the work.

Chapter 3 has been published in *Metabolic Engineering*, Volume 49, September 2018, under doi 10.1016/j.ymben.2018.07.003. I performed the feeding experiments with the assistance of Longyun Guo, obtained the experimental data. Longyun Guo conducted the computational simulation. Longyun Guo wrote the main body of the manuscript and I wrote part of the Methods. Longyun Guo and I are both co-first authors for this manuscript.

Chapter 4 will be published in the near future. I prepared the plant materials for RNA sequencing at Purdue Genomics Core Facility and analyzed the sequencing data on Purdue Community Cluster Snyder run by Purdue Information technology Research Computing. I also conducted the metabolite profiling work.

ABSTRACT

Author: Wang, Peng. PhD

Institution: Purdue University

Degree Received: May 2019

Title: Measuring and Modeling of Phenylpropanoid Metabolic Flux in Arabidopsis

Major Professor: Clint Chapple

Plants naturally deposit a significant amount of carbon towards lignin, a polymer that imparts mechanical strength to cell walls but impedes our utilization of the polysaccharides in lignocellulosic biomass. Genetic engineering of lignin has demonstrated profound success in improving the processing of the biomass. Lignin is derived from the phenylpropanoid pathway, the architecture of which is well understood based upon the biochemical and genetic studies conducted to date. In contrast, we lack a systematic and quantitative view of the factors that determine carbon flux into and within this branched metabolic pathway in plants. To explore the control of carbon allocation for phenylalanine and lignin biosynthesis, we have developed a kinetic model of the pathway in Arabidopsis to test the regulatory role of several key enzymatic steps. We first established a ^{13}C isotope feeding system for the measurement of flux using excised wild-type Arabidopsis stems. The excised stems continued to grow and lignify in our feeding system. When ring $^{13}\text{C}_6$ -labeled phenylalanine ($[^{13}\text{C}_6]$ -Phe) was supplied to excised stems, isotope label was rapidly incorporated into soluble intermediates and lignin. Using this approach, we then analyzed metabolite pool sizes and isotope abundances of the pathway intermediates in a time course from stems fed with $[^{13}\text{C}_6]$ -Phe of different concentrations, and used these data to parameterize a kinetic model constructed with Michaelis-Menten kinetics. Our model of the general phenylpropanoid pathway captured the dynamic trends of metabolite pools *in vivo* and predicted the metabolic profiles of an independent feeding experiment. Based on the model simulation, we found that subcellular sequestration of pathway intermediates is necessary to maintain lignification homeostasis when metabolites are over-accumulated. Both the measurements and simulation suggested that the availability of substrate Phe is one limiting factor for lignin flux in developing stems. This finding indicates new gene targets for lignin manipulation in plants. To extend our kinetic model to simulate flux distribution in response to genetic perturbations, we conducted an RNA-sequencing experiment in wild type and 13 plants with modified lignification, and integrated

the transcriptional data with the metabolic profiles. We found that the biosynthesis of Phe and lignification are tightly coordinated at transcriptional level. The coregulation of the shikimate and phenylpropanoid pathways involves transcriptional and post-translational regulatory mechanisms to maintain pathway homeostasis. Our results also indicate that induction of Phe supply and enhancement of PAL activity are both effective strategies to increase carbon flux into the phenylpropanoid network.

In this interdisciplinary project, we have taken various system biology approaches to understand metabolic flux towards lignin, the second most abundant carbon sink in nature. We have combined isotope labeling aided flux measurements and mathematical simulation, and have integrated metabolome data with transcriptome profiles. The experiments and analysis have been conducted in both wild-type *Arabidopsis* and those with perturbed lignification. The novel work not only provides insight into our knowledge of phenylpropanoid metabolism, but also creates a framework to systematically assemble gene expression, enzyme activity, and metabolite accumulation to study metabolic fluxes, the ultimate functional phenotypes of biochemical networks.

CHAPTER 1. GENETIC MANIPULATION OF LIGNOCELLULOSIC BIOMASS FOR BIOENERGY

1.1 Abstract

Lignocellulosic biomass represents an abundant and sustainable raw material for biofuel production. The recalcitrance of biomass to degradation increases the estimated cost of biofuel production and limits its competitiveness in the market. Genetic engineering of lignin, a major recalcitrance factor, improves saccharification and thus the potential yield of biofuels. Recently, our understanding of lignification and its regulation has been advanced by new studies in various systems, all of which further enhances our ability to manipulate the biosynthesis and deposition of lignin in energy crops for producing cost-effective second generation biofuels.

1.2 Introduction

Lignocellulose accounts for 70% of the carbon fixed by photosynthesis in land plants every year and a substantial amount of this biomass is available for use by humans [1]. Conversion of lignocellulosic feedstocks to biofuels has become increasingly attractive as a partial replacement for petroleum [2,3]. In contrast to bioethanol derived from starch, proposed lignocellulosic biofuel production facilities would use forest or crop residues and perennial grasses as feedstocks [1,2]. Biochemical and genetic investigations of lignocellulose in energy crops such as poplar, sorghum, sugarcane, miscanthus, and switchgrass are motivated by improving biofuel yield and providing a carbon-neutral source of transportation fuel.

Plant secondary cell walls are largely composed of cellulose, hemicelluloses and lignin. Cellulose, a β -1,4-linked polymer of glucose, is the most abundant component in secondary cell walls and hence the major source of sugar units used in biofuel production. Hemicelluloses include mannans, xylans, xyloglucans, or mixed-linkage glucans, which interact with cellulose to form a semi-rigid matrix [4]. Lignin is a heterogeneous polymer predominantly constituted of three subunits, namely *p*-hydroxyphenyl (H), guaiacyl (G), and syringyl (S) monolignols. These monomers are derived from phenylalanine through the phenylpropanoid pathway, the structure of which is now well established (Figure 1.1) [5]. Lignin imparts mechanical strength and hydrophobicity to cell walls, but decreases saccharification efficiency and thereby increases the

cost of lignocellulosic biofuel production. Fortunately, genetic modification of lignification results in changes in lignin content and/or composition, which in many cases leads to improved enzymatic digestion of the polysaccharide components of the cell wall [6-8]. Here, we summarize recent advances in the genetic manipulation of model plants and energy crops for enhanced fermentable sugar release.

1.3 Identification of new genes for lignin manipulation

Lignification is the consequence of coordinated monolignol biosynthesis and oxidative polymerization, both of which are tightly regulated transcriptionally and post-translationally [5,9]. The identification and functional characterization of the genes involved in these processes provide potential targets for modifying lignin and improving saccharification. For example, the recent discovery of caffeoyl shikimate esterase (CSE) revealed that this enzyme catalyzes the release of caffeate from caffeoyl shikimate which can then be utilized by 4-coumarate CoA ligase (4CL) to form caffeoyl CoA, an important intermediate for monolignol biosynthesis (Figure 1.1) [10]. In effect, the CSE/4CL route bypasses the transfer of the caffeoyl moiety from caffeoyl shikimate back to CoA by hydroxycinnamoyl CoA: shikimate hydroxycinnamoyl transferase (HCT). The *cse* mutant accumulates less total lignin but over 30-fold more H units compared with wild-type *Arabidopsis*. Although the *cse* mutant produces less biomass, it releases 75% more glucose on a per plant basis during saccharification without pretreatment. As another example, cytochrome P450 reductases (CPRs) are also important for lignification because they transfer electrons from NADPH to the three cytochrome P450 hydroxylases required for monolignol synthesis, cinnamate 4-hydroxylase (C4H), *p*-coumarate 3'-hydroxylase (C3'H), and ferulate 5-hydroxylase (F5H) [11]. Although it had been assumed that the CPRs in *Arabidopsis* were functionally redundant, it was recently demonstrated that the *Arabidopsis atr2* mutant, defective in CPR2, produces less total lignin with a 10-fold enhancement in H unit content, and also exhibits improved glucose yields when alkaline-pretreated stems are saccharified [12].

Disruption of genes required for lignin polymerization also results in reduced lignin content and altered composition. Laccases (LACs) has been proposed to activate monolignols for radical coupling in the apoplast [5]. The first *in vivo* evidence came from the *Arabidopsis lac4 lac17* double mutant, which displays an irregular xylem phenotype and has 40% less lignin with an increased S/G ratio. These phenotypes led to higher glucose release when the mutant's cell wall

residue was hydrolyzed with commercial cellulase [13]. The *Brachypodium lac5* mutant also deposits less lignin and exhibits enhanced saccharification, further demonstrating that *LACs* are potential genetic targets for improving biofuel production in monocot energy crops [14]. Like *LACs*, peroxidases have been shown to oxidize monolignols [5] and participate in lignin deposition in the Casparian strip [15], but more *in planta* data are necessary to define the biological functions of different peroxidases in vascular lignification [16,17]. Clearly, enzymes involved in both monomer synthesis and polymerization hold promise as tools for the improvement of biofuel crops.

The introduction of genes from other organisms is an alternative strategy to perturb lignin biosynthesis that is independent of the modification of endogenous plant genes. For example, the plastidial expression of 3-dehydroshikimate dehydratase from *Corynebacterium glutamicum* in *Arabidopsis* diverts flux from the shikimate pathway to synthesize protocatechuate, causes a 50% reduction in total lignin content and results in enhanced sugar release after enzymatic hydrolysis [18]. An interesting observation is that diverting carbon flux away from the shikimate pathway affects downstream lignification but not the pools of aromatic amino acids or salicylic acid [18]. These findings demonstrate that expression of foreign genes can also be used to redirect flux away from lignin and to reduce the recalcitrance of biomass for bioenergy production.

1.4 Modifications of lignin monomers

In addition to the three primary *p*-hydroxycinnamyl alcohols, native lignins also contain numerous additional phenolic metabolites such as hydroxycinnamaldehydes, hydroxycinnamic acids, and hydroxybenzaldehydes [19]. When lignin biosynthesis is genetically perturbed, phenylpropanoid intermediates that accumulate as a result can be incorporated into the polymer (Figure 1.2) [5-7]. Indeed, overexpression of F5H in caffeic acid *O*-methyltransferase (COMT) - deficient *Arabidopsis* mutant resulted in an increased deposition of 5-hydroxyguaiacyl units in lignin (Figure 1.1) [20,21]. Similarly, a catechyl lignin polymer derived from caffeyl alcohol was identified in *CAFFEOYL CoA O-METHYLTRANSFERASE (CCoAOMT)* suppressed pine cell culture (Figure 1.1) like that recently characterized in the seed coat of a variety of members of the Orchidaceae, Cactaceae, Euphorbiaceae, and Cleomaceae [22-24]. The abundance of benzodioxane structures in these two novel lignin types can minimize potential cross-linking between lignin α -carbons and hemicellulosic alcohol or acid groups (Figure 1.3) [19]. This

suggests that the engineering of plants that deposit catechyl lignin might lead to improved cellulose conversion, similar to the COMT-deficient switchgrass and alfalfa that have been developed recently that deposit lignin with increased 5-hydroxyconiferyl alcohol monomer content [25-27].

The plasticity of lignin biosynthesis and polymerization makes it possible to incorporate new monomers and thus new functionality into the lignin polymer or to interfere with polymerization. An array of phenolic compounds has been proposed as promising candidates to improve saccharification through these mechanisms [19]. The first example successfully realized in vivo involved heterologous expression of an *Angelica sinensis* feruloyl CoA monolignol transferase in poplar, which resulted in the incorporation of ferulate conjugates into the lignin backbone (Figure 1.3) [28]. The novel lignins formed did not impact the growth of the poplar trees but the ester bonds within these novel lignin building blocks could be easily cleaved by alkaline pretreatment post-harvest, thus enhancing the release of sugar [28]. Instead of “editing” monolignols before polymerization, enzymatic modification of the lignin polymer itself can also result in readily cleavable bonds. A C- α dehydrogenase from *Sphingobium* was shown to oxidize the hydroxyl group of the α -carbon in β -O-4 linked dilignols and oligolignols in vitro or when expressed in Arabidopsis (Figure 1.3) [29]. Lignin of the transgenic lines showed only a small increase in α -keto β -O-4 linkages presumably because most of the C- α dehydrogenase appeared to localize to the cytoplasm even though it had been targeted for secretion [29]. It might be possible to attain higher levels of α -keto β -O-4 linkages in lignin in the future if the C- α dehydrogenase can be successfully targeted to the apoplast of lignifying cells.

The elongation of the lignin polymer is mainly the result of β -O-4 coupling between the β -carbon of monomers and the 4-O position of oligomers [7,19]. Monomers that lack a β -carbon or in which the para-hydroxyl group is blocked prevent the polymer from growing. For example, hydroxybenzaldehydes are known as minor non-conventional monomers in a variety of species and form end groups in natural lignins (Figure 1.3) [30]. To increase these end units and alter lignin architecture, a *Pseudomonas fluorescens* hydroxycinnamoyl CoA hydratase-lyase was expressed in Arabidopsis [31]. Although the transgenics displayed normal growth, they contained more reduced-length lignins and provided an increased yield of sugars after various pretreatments [31]. Methylation of the para-hydroxyl group of monolignols is another way to reduce polymerization by preventing β -O-4 linkages with the free monomers. The *Clarkia breweri* isoeugenol O-methyltransferase was engineered to catalyze the 4-O-methylation of monolignols,

and expression of this enzyme in *Arabidopsis* reduced lignin content and led to elevated glucose yield without pretreatment [32]. Taken together, these results show that genetic manipulation of lignification can result in the incorporation of non-conventional monomers with positive impacts on biomass utility.

1.5 Regulations of lignocellulose biosynthesis

A hierarchical transcriptional network regulates secondary cell wall biosynthesis in plants, and coordinates cell elongation, differentiation, synthesis of polysaccharides and lignin, and programmed cell death [9]. The top-level master switches include a series of tissue-specific NAM, ATAF1,2, and CUC2 (NAC) transcription factors, which activate the secondary-level MYB master switches, such as MYB46 and MYB83 in *Arabidopsis*. These master switches together turn on downstream transcription factors and directly regulate the biosynthetic genes of the secondary cell wall [9,33]. Feed-forward loops between different transcription factors make the network robust in regulating secondary cell wall synthesis. To describe this network, Taylor-Teeples *et al.* utilized DNA-protein binding assays to study the complex regulatory feed-forward cascades controlling *Arabidopsis* lignocellulosic biosynthesis [34]. In terms of feedstock genetic engineering, one advantage of transcription factor manipulation is that multiple genes can be simultaneously regulated to restrict or enhance flux towards lignocellulose. Such an approach avoids the necessity of stacking of biosynthetic genes (or their mutants) to achieve desired phenotypes. For example, overexpression of *MYB4* in switchgrass, a transcriptional repressor of lignification, reduces the transcription of 10 lignin biosynthetic genes, decreases lignin content and increases cellulosic ethanol yield up to 2.6 fold [35,36].

It was recently demonstrated that transcriptional regulation of phenylpropanoid metabolism in plants is also dependent on the Mediator complex. Subunits of Mediator, MED5a and MED5b, suppress the steady state level of phenylpropanoid biosynthetic mRNAs in *Arabidopsis* [37]. Knocking-out *MED5a* and *MED5b* enhances phenylpropanoid gene expression and end product accumulation. Surprisingly, it also substantially alleviates the mis-regulation of transcription in the stunted and lignin-deficient *Arabidopsis c3 h* mutant, in which one-third of the transcriptome is differentially expressed, rescues its arrested growth and restores lignin deposition to wild-type levels (Figure 1.4) [38,39]. Most importantly, due to the defect in *C3 H*, the mutant's

lignin is almost exclusively derived from H subunits, demonstrating that these monomers can be polymerized into a functional lignin that supports normal growth. The deposition of this novel lignin doubles the saccharification efficiency of the mutant's biomass compared to the wild-type [38].

In addition to regulation at the transcriptional level, posttranslational regulation also controls lignin biosynthesis [40]. It has recently been shown that ubiquitination and degradation of phenylalanine ammonia lyase (PAL), the first enzyme in the phenylpropanoid pathway, is mediated by Kelch-repeat-containing F box (KFB) proteins, components of the SCF type E3 ligase [41,42]. Overexpression of the KFBs facilitates the proteolysis of PAL and results in a reduction in lignin content and soluble phenylpropanoids in *Arabidopsis* [41,42] and may provide a novel approach to biomass modification in the future.

1.6 Engineering of energy crops for lignocellulosic utilization

An increasing number of studies conducted directly on energy crops have identified promising strains for biofuel production, including spontaneous or chemically-induced mutants. Mutant screens are particularly facile when visible markers are available. Brown midrib mutations, first found in natural maize mutants, are clearly visible to the unaided eye, and are associated with reduced lignin content or altered lignin composition [43]. Random mutagenesis screens of sorghum were conducted to identify brown midrib (*bmr*) mutants and so far three independent loci have been isolated and found to encode 4CL (*bmr2*), CAD (*bmr6*), COMT (*bmr12*), respectively [43-45]. Characterization of these mutants showed a 7-20% increase in saccharification efficiency and a further 20% increase when *bmr6* and *bmr12* mutations are stacked [45,46].

Translation of genetic engineering strategies established in model plants into energy crops requires a comprehensive knowledge of lignification in these target species. Advanced sequencing technologies have enabled the systematic identification of lignin biosynthetic candidate genes in poplar (*Populus trichocarpa*), *Eucalyptus grandis*, switchgrass (*Panicum virgatum*), and sugarcane (*Saccharum spp.*) thus facilitating reverse genetic analysis in these crops [47-50]. For example, RNAi mediated down-regulation of *C3'H* in poplar resulted in lower lignin content and higher glucose release by enzymatic hydrolysis (Figure 1.2) [51,52] and the CRISPR/Cas9 system was successfully employed to mutate *4CL1* and *4CL2* in hybrid poplar [53].

To be economically viable for biofuel production, modified crops will need to thrive in the field and yield high amounts of biomass. A five-year field trial study showed a significant increase in enzymatic hydrolysis of cellulose in transgenic poplar deficient in CCoAOMT but no impacts on its growth or biomass [54]. Similarly, field-grown *COMT*-downregulated or MYB4 overexpressing switchgrass yielded increased amounts of fermentable sugars and comparable amounts of dry biomass [55,56]. Importantly, in the field these transgenic plants displayed no differences in disease susceptibility compared to wild-type controls. Elevated saccharification of lignocellulose was also observed in *COMT*-suppressed sugarcane and *CCR* downregulated poplar in field trials, but the improvements were accompanied by a reduction in biomass [57,58]. Although the higher cellulose conversion in the transgenic sugarcane and poplar compensates for the reduced biomass, the overall yield penalty in these lines remains an obstacle to be overcome before industrial utilization.

1.7 Conclusions

The production of cost-effective biofuels from abundant and renewable lignocellulosic feedstocks is impeded by the inherent recalcitrance of the biomass. Various genetic manipulations of lignification have demonstrated the power to reduce this stumbling block and increase the yield of fermentable sugars from engineered biomass. Alteration of lignin content and composition can be achieved by modification of endogenous plant genes or expression of heterologous genes. The introduction of novel monomers and increased levels of non-conventional subunits result in lignin with new functional groups and ease the removal of lignin for saccharification. Further, regulation of lignification at both transcriptional and post-translational levels is shown to be effective targets to engineer lignin. Extension of the established strategies from model systems into energy crops and assessment of the mutants in the field are necessary for successful industrial application. Future studies must focus on overcoming the yield loss that often accompanies lignin modification.

1.8 References

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest

•• of outstanding interest

1. Pauly M, Keegstra K: Cell-wall carbohydrates and their modification as a resource for biofuels. *Plant J* 2008, 54:559-568.
2. Somerville C: Next generation biofuels. *AIP Conf Proc* 2015, 1652:44-50.
3. Limayem A, Ricke SC: Lignocellulosic biomass for bioethanol production: current perspectives, potential issues and future prospects. *Prog Energy Combust Sci* 2012, 38:449-467.
4. Pauly M, Gille S, Liu LF, Mansoori N, de Souza A, Schultink A, Xiong GY: Hemicellulose biosynthesis. *Planta* 2013, 238:627-642.
5. Bonawitz ND, Chapple C: The genetics of lignin biosynthesis: connecting genotype to phenotype. *Annu Rev Genet* 2010, 44:337-363.
6. Liu CJ, Cai Y, Zhang Y, Gou M, CJ, Cai Y, Zhang X, Gou M, Yang H. Tailoring lignin biosynthesis for efficient and sustainable biofuel production. *Plant Biotechnol J* 2014, 12:1154-1162.
- 7. Van Acker R, Vanholme R, Storme V, Mortimer JC, Dupree P, Boerjan W: Lignin biosynthesis perturbations affect secondary cell wall composition and saccharification yield in *Arabidopsis thaliana*. *Biotechnol Biofuels* 2013, 6.
Describes comprehensively the effect of perturbation at each lignin biosynthetic step on secondary cell wall composition and saccharification efficiency.
8. Li X, Ximenes E, Kim Y, Slininger M, Meilan R, Ladisch M, Chapple C: Lignin monomer composition affects *Arabidopsis* cell-wall degradability after liquid hot water pretreatment. *Biotechnol Biofuels* 2010, 3:27.
9. Zhong R, Ye ZH: Secondary cell walls: biosynthesis, patterned deposition and transcriptional regulation. *Plant Cell Physiol* 2015, 56:195-214.
- 10. Vanholme R, Cesarino I, Rataj K, Xiao Y, Sundin L, Goeminne G, Kim H, Cross J, Morreel K, Araujo P, et al.: Caffeoyl shikimate esterase (CSE) is an enzyme in the lignin biosynthetic pathway in *Arabidopsis*. *Science* 2013, 341:1103-1106.
Reports the identification of a new enzyme involved in the phenylpropanoid metabolism and causes a revision of the pathway.
11. Urban P, Mignotte C, Kazmaier M, Delorme F, Pompon D: Cloning, yeast expression, and characterization of the coupling of two distantly related *Arabidopsis thaliana* NADPH-cytochrome P450 reductases with P450 CYP73A5. *J Biol Chem* 1997, 272:19176-19186.

12. Sundin L, Vanholme R, Geerinck J, Goeminne G, Hofer R, Kim H, Ralph J, Boerjan W: Mutation of the inducible *ARABIDOPSIS THALIANA* *CYTOCHROME P450 REDUCTASE2* alters lignin composition and improves saccharification. *Plant Physiol* 2014, 166:1956-1971.
13. Berthet S, Demont-Caulet N, Pollet B, Bidzinski P, Cezard L, Le Bris P, Borrega N, Herve J, Blondet E, Balzergue S, et al.: Disruption of *LACCASE4* and *17* results in tissue-specific alterations to lignification of *Arabidopsis thaliana* stems. *Plant Cell* 2011, 23:1124-1137.
14. Wang Y, Bouchabke-Coussa O, Le Bris P, Antelme S, Soulhat C, Gineau E, Dalmais M, Bendahmane A, Morin H, Mouille G, et al.: *LACCASE5* is required for lignification of the *Brachypodium distachyon* culm. *Plant Physiol* 2015, 168:192-204.
15. Lee Y, Rubio MC, Alassimone J, Geldner N: A mechanism for localized lignin deposition in the endodermis. *Cell* 2013, 153:402-412.
16. Francoz E, Ranocha P, Nguyen-Kim H, Jamet E, Burlat V, Dunand C: Roles of cell wall peroxidases in plant development. *Phytochem* 2015, 112:15-21.
17. Herrero J, Esteban-Carrasco A, Zapata JM: Looking for *Arabidopsis thaliana* peroxidases involved in lignin biosynthesis. *Plant Physiol Biochem* 2013, 67:77-86.
- 18. Eudes A, Sathitsuksanoh N, Baidoo EE, George A, Liang Y, Yang F, Singh S, Keasling JD, Simmons BA, Loque D: Expression of a bacterial 3-dehydroshikimate dehydratase reduces lignin content and improves biomass saccharification efficiency. *Plant Biotechnol J* 2015, <http://dx.doi.org/10.1111/pbi.12310>.
Reports that the manipulation of the shikimate pathway diverts the flux away from lignin and leads to increased enzymatic digestibility of lignocellulose.
- 19. Vanholme R, Morreel K, Darrah C, Oyarce P, Grabber JH, Ralph J, Boerjan W: Metabolic engineering of novel lignin in biomass crops. *New Phytol* 2012, 196:978-1000.
Describes in-depth a library of candidate phenolic compounds as lignin monomers for improvement in lignocellulosic biofuel production.
20. Weng JK, Mo H, Chapple C: Over-expression of F5H in COMT-deficient *Arabidopsis* leads to enrichment of an unusual lignin and disruption of pollen wall formation. *Plant J* 2010, 64:898-911.

21. Vanholme R, Ralph J, Akiyama T, Lu F, Pazo JR, Kim H, Christensen JH, Van Reusel B, Storme V, De Rycke R, et al.: Engineering traditional monolignols out of lignin by concomitant up-regulation of F5H1 and down-regulation of COMT in Arabidopsis. *Plant J* 2010, 64:885-897.
- 22. Chen F, Tobimatsu Y, Havkin-Frenkel D, Dixon RA, Ralph J: A polymer of caffeyl alcohol in plant seeds. *Proc Natl Acad Sci USA* 2012, 109:1772-1777.
Reports the first observation of natural catechyl lignin in plant seeds.
23. Tobimatsu Y, Chen F, Nakashima J, Escamilla-Trevino LL, Jackson L, Dixon RA, Ralph J: Coexistence but independent biosynthesis of catechyl and guaiacyl/syringyl lignin polymers in seed coats. *Plant Cell* 2013, 25:2587-2600.
24. Wagner A, Tobimatsu Y, Phillips L, Flint H, Torr K, Donaldson L, Pears L, Ralph J: CCoAOMT suppression modifies lignin composition in *Pinus radiata*. *Plant J* 2011, 67:119-129.
25. Fu C, Mielenz JR, Xiao X, Ge Y, Hamilton CY, Rodriguez M, Jr., Chen F, Foston M, Ragauskas A, Bouton J, et al.: Genetic manipulation of lignin reduces recalcitrance and improves ethanol production from switchgrass. *Proc Natl Acad Sci USA* 2011, 108:3803-3808.
26. Dien BS, Miller DJ, Hector RE, Dixon RA, Chen F, McCaslin M, Reisen P, Sarath G, Cotta MA: Enhancing alfalfa conversion efficiencies for sugar recovery and ethanol production by altering lignin composition. *Bioresour Technol* 2011, 102:6479-6486.
27. Sattler SE, Palmer NA, Saballos A, Greene AM, Xin ZG, Sarath G, Vermerris W, Pedersen JF: Identification and characterization of four missense mutations in brown midrib 12 (Bmr12), the caffeic O-methyltransferase (COMT) of sorghum. *BioEnergy Res* 2012, 5:855-865.
- 28. Wilkerson CG, Mansfield SD, Lu F, Withers S, Park JY, Karlen SD, Gonzales-Vigil E, Padmakshan D, Unda F, Rencoret J, et al.: Monolignol ferulate transferase introduces chemically labile linkages into the lignin backbone. *Science* 2014, 344:90-93.
Demonstrates that monolignol ferulate can be synthesized and incorporated into lignin polymer and result in readily cleavable ester bonds in transgenic poplars expressing a monolignol ferulate transferase.

29. Tsuji Y, Vanholme R, Tobimatsu Y, Ishikawa Y, Foster CE, Kamimura N, Hishiyama S, Hashimoto S, Shino A, Hara H, et al.: Introduction of chemically labile substructures into Arabidopsis lignin through the use of LigD, the Ca-dehydrogenase from *Sphingobium* sp. strain SYK-6. *Plant Biotechnol J* 2015, <http://dx.doi.org/10.1111/pbi.12316>.
30. Kim H, Ralph J, Lu FC, Ralph SA, Boudet AM, MacKay JJ, Sederoff RR, Ito T, Kawai S, Ohashi H, et al.: NMR analysis of lignins in CAD-deficient plants. Part 1. Incorporation of hydroxycinnamaldehydes and hydroxybenzaldehydes into lignins. *Org Biomol Chem* 2003, 1:268-281.
31. Eudes A, George A, Mukerjee P, Kim JS, Pollet B, Benke PI, Yang F, Mitra P, Sun L, Cetinkol OP, et al.: Biosynthesis and incorporation of side-chain-truncated lignin monomers to reduce lignin polymerization and enhance saccharification. *Plant Biotechnol J* 2012, 10:609-620.
32. Zhang K, Bhuiya MW, Pazo JR, Miao Y, Kim H, Ralph J, Liu CJ: An engineered monolignol 4-*O*-methyltransferase depresses lignin biosynthesis and confers novel metabolic capability in Arabidopsis. *Plant Cell* 2012, 24:3135-3152.
33. Zhong RQ, McCarthy RL, Lee C, Ye ZH: Dissection of the transcriptional program regulating secondary wall biosynthesis during wood formation in poplar. *Plant Physiol* 2011, 157:1452-1468.
- 34. Taylor-Teeple M, Lin L, de Lucas M, Turco G, Toal TW, Gaudinier A, Young NF, Trabucco GM, Veling MT, Lamothe R, et al.: An Arabidopsis gene regulatory network for secondary cell wall synthesis. *Nature* 2015, 517:571-U307.
Reports a comprehensive protein-DNA regulatory network for lignocellulose deposition in Arabidopsis root and stem, supporting the feed-forward loops for robust regulation of secondary cell wall synthesis.
35. Shen H, Poovaiah CR, Ziebell A, Tschaplinski TJ, Pattathil S, Gjersing E, Engle NL, Katahira R, Pu Y, Sykes R, et al.: Enhanced characteristics of genetically modified switchgrass (*Panicum virgatum* L.) for high biofuel production. *Biotechnol Biofuels* 2013, 6:71.
36. Shen H, He XZ, Poovaiah CR, Wuddineh WA, Ma JY, Mann DGJ, Wang HZ, Jackson L, Tang YH, Stewart CN, et al.: Functional characterization of the switchgrass (*Panicum virgatum*) R2R3-MYB transcription factor PvMYB4 for improvement of lignocellulosic feedstocks. *New Phytol* 2012, 193:121-136.

37. Bonawitz ND, Soltau WL, Blatchley MR, Powers BL, Hurlock AK, Seals LA, Weng JK, Stout J, Chapple C: REF4 and RFR1, subunits of the transcriptional coregulatory complex mediator, are required for phenylpropanoid homeostasis in Arabidopsis. *J Biol Chem* 2012, 287:5434-5445.
- 38. Bonawitz ND, Kim JI, Tobimatsu Y, Ciesielski PN, Anderson NA, Ximenes E, Maeda J, Ralph J, Donohoe BS, Ladisch M, Chapple C: Disruption of Mediator rescues the stunted growth of a lignin-deficient Arabidopsis mutant. *Nature* 2014, 509:376-380.
Reports that Mediator is required for the dwarfism and lignin-deficiency in Arabidopsis *ref8* mutants, and that H lignin can support the normal growth of plants and reduce the recalcitrance of lignocellulosic biomass.
39. Franke R, Hemm MR, Denault JW, Ruegger MO, Humphreys JM, Chapple C: Changes in secondary metabolism and deposition of an unusual lignin in the *ref8* mutant of Arabidopsis. *Plant J* 2002, 30:47-59.
40. Zhang XB, Liu CJ: Multifaceted regulations of gateway enzyme phenylalanine ammonia-lyase in the biosynthesis of phenylpropanoids. *Mol Plant* 2015, 8:17-27.
- 41. Zhang X, Gou M, Liu CJ: Arabidopsis Kelch repeat F-box proteins regulate phenylpropanoid biosynthesis via controlling the turnover of phenylalanine ammonia-lyase. *Plant Cell* 2013, 25:4994-5010.
Reports that KFB proteins mediate the degradation of PAL and regulate phenylpropanoid metabolism at the post-translational level.
42. Zhang XB, Gou MY, Guo CR, Yang HJ, Liu CJ: Down-regulation of Kelch domain-containing F-box protein in Arabidopsis enhances the production of (poly)phenols and tolerance to ultraviolet radiation. *Plant Physiol* 2015, 167:337-U548.
43. Sattler SE, Funnell-Harris DL, Pedersen JF: Brown midrib mutations and their importance to the utilization of maize, sorghum, and pearl millet lignocellulosic tissues. *Plant Science* 2010, 178:229-238.
44. Xin Z, Wang ML, Barkley NA, Burow G, Franks C, Pederson G, Burke J. Applying genotyping (TILLING) and phenotyping analyses to elucidate gene function in a chemically induced sorghum mutant population. *BMC Plant Biol* 2008, 8:103.

45. Saballos A, Vermerris W, Rivera L, Ejeta G: Allelic association, chemical characterization and saccharification properties of *brown midrib* mutants of sorghum (*Sorghum bicolor* (L.) Moench). *BioEnergy Res* 2008, 1:193-204.
46. Dien BS, Sarath G, Pedersen JF, Sattler SE, Chen H, Funnell-Harris DL, Nichols NN, Cotta MA: Improved sugar conversion and ethanol yield for forage sorghum (*Sorghum bicolor* L. Moench) lines with reduced lignin contents. *BioEnergy Res* 2009, 2:153-164.
47. Carocha V, Soler M, Hefer C, Cassan-Wang H, Fevereiro P, Myburg AA, Paiva JA, Grima-Pettenati J: Genome-wide analysis of the lignin toolbox of *Eucalyptus grandis*. *New Phytol* 2015, 206:1297-1313.
48. Bottcher A, Cesarino I, dos Santos AB, Vicentini R, Mayer JLS, Vanholme R, Morreel K, Goeminne G, Moura JCMS, Nobile PM, et al.: Lignification in sugarcane: biochemical characterization, gene discovery, and expression analysis in two genotypes contrasting for lignin content. *Plant Physiol* 2013, 163:1539-1557.
49. Shen H, Mazarei M, Hisano H, Escamilla-Trevino L, Fu C, Pu Y, Rudis MR, Tang Y, Xiao X, Jackson L, et al.: A genomics approach to deciphering lignin biosynthesis in switchgrass. *Plant Cell* 2013, 25:4342-4361.
50. Shi R, Sun YH, Li Q, Heber S, Sederoff R, Chiang VL: Towards a systems approach for lignin biosynthesis in *Populus trichocarpa*: transcript abundance and specificity of the monolignol biosynthetic genes. *Plant Cell Physiol* 2010, 51:144-163.
51. Mansfield SD, Kang KY, Chapple C: Designed for deconstruction - poplar trees altered in cell wall lignification improve the efficacy of bioethanol production. *New Phytol* 2012, 194:91-101.
52. Coleman HD, Park JY, Nair R, Chapple C, Mansfield SD: RNAi-mediated suppression of *p*-coumaroyl-CoA 3'-hydroxylase in hybrid poplar impacts lignin deposition and soluble secondary metabolism. *Proc Natl Acad Sci USA* 2008, 105:4501-4506.
- 53. Zhou X, Jacobs TB, Xue LJ, Harding SA, Tsai CJ: Exploiting SNPs for biallelic CRISPR mutations in the outcrossing woody perennial *Populus* reveals 4-coumarate:CoA ligase specificity and redundancy. *New Phytol* 2015, <http://dx.doi.org/10.1111/nph.13470>.
Reports the first utilization of CRISPR/Cas9 system to manipulate lignin biosynthetic genes in plants.

54. Wang HZ, Xue YX, Chen YJ, Li RF, Wei JH: Lignin modification improves the biofuel production potential in transgenic *Populus tomentosa*. *Ind Crops Prod* 2012, 37:170-177.
55. Baxter HL, Mazarei M, Labbe N, Kline LM, Cheng QK, Windham MT, Mann DGJ, Fu CX, Ziebell A, Sykes RW, et al.: Two-year field analysis of reduced recalcitrance transgenic switchgrass. *Plant Biotechnol J* 2014, 12:914-924.
56. Baxter HL, Poovaiah C, Yee K, Mazarei M, Rodriguez M, Jr., Thompson O, Shen H, Turner G, Decker S, Sykes R, et al.: Field evaluation of transgenic switchgrass plants overexpressing *PvMYB4* for reduced biomass recalcitrance. *BioEnergy Res* 2015:1-12.
57. Van Acker R, Leple JC, Aerts D, Storme V, Goeminne G, Ivens B, Legee F, Lapierre C, Piens K, Van Montagu MCE, et al.: Improved saccharification and ethanol yield from field-grown transgenic poplar deficient in cinnamoyl-CoA reductase. *Proc Natl Acad Sci USA* 2014, 111:845-850.
58. Jung JH, Vermerris W, Gallo M, Fedenko JR, Erickson JE, Altpeter F: RNA interference suppression of lignin biosynthesis increases fermentable sugar yields for biofuel production from field-grown sugarcane. *Plant Biotechnol J* 2013, 11:709-716.

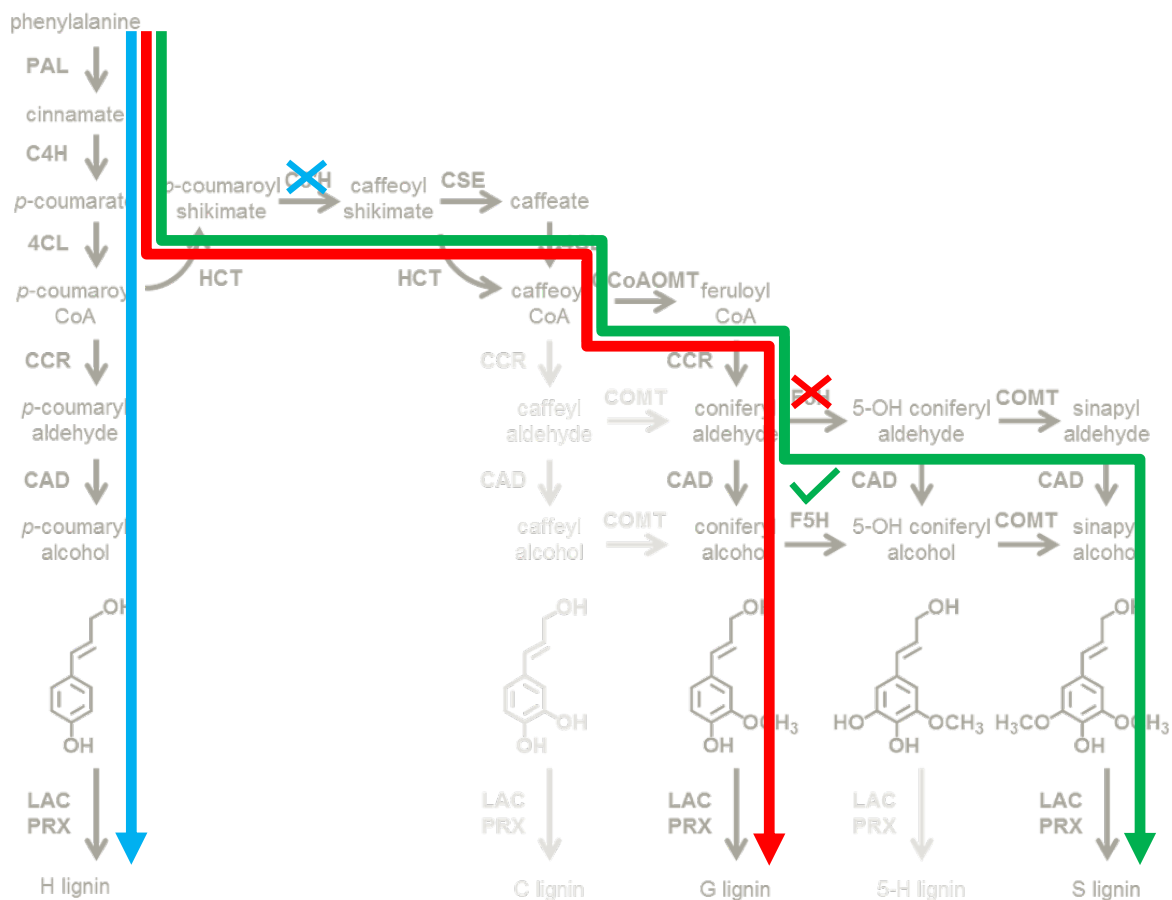


Figure 1.2 A simplified model illustrating that manipulations of lignin biosynthetic genes change pathway flux.

In flowering plants, the pathway generates H, G, and S lignin. When *C3'H* is perturbed (blue X), the flux towards G and S is blocked and mutants deposit H lignin (blue line). When *F5H* is silenced (red X), the S branch is blocked and mutants contain almost exclusively G lignin (red line). When *F5H* is overexpressed (green check), the G flux is redirected towards S monomers so transgenics accumulate predominantly S lignin (green line).

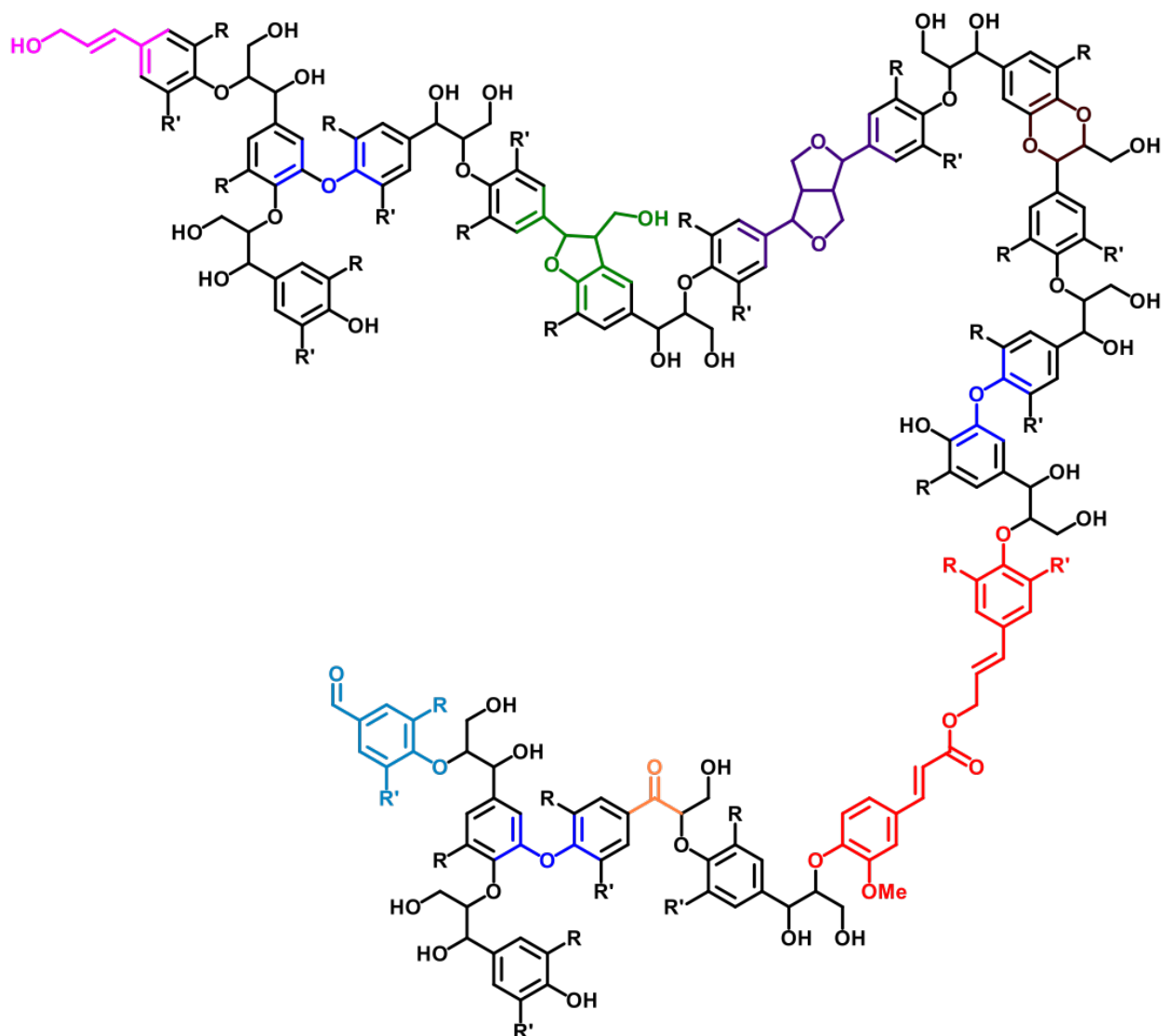


Figure 1.3 A model of lignin polymer illustrates the native and modified monomers and structural units.

The major monomers are coniferyl alcohols ($R=OMe$, $R'=H$) and sinapyl alcohols ($R=R'=OMe$), and the major structural units are β -ether units (black) in wild-type plants. Wild-type lignin also includes phenylcoumaran units (green), resinol units (purple), and biphenyl ether units (blue).

Cinnamyl alcohols ($R=OMe$, $R'=OMe$ or H) (pink) and hydroxybenzaldehydes ($R=OMe$, $R'=OMe$ or H) (cyan) are often seen as end groups. Incorporation of caffeyl alcohol ($R=H$) or 5-hydroxyconiferyl alcohol ($R=OMe$) results in benzodioxane units (brown) in the lignin polymers. Incorporation of monolignol ferulate (red) introduces the ester bond into the lignin backbone, forming ZipLigninTM. Oxidation of hydroxyl group on the α -carbon results in α -keto β -O-4 linkages (orange).

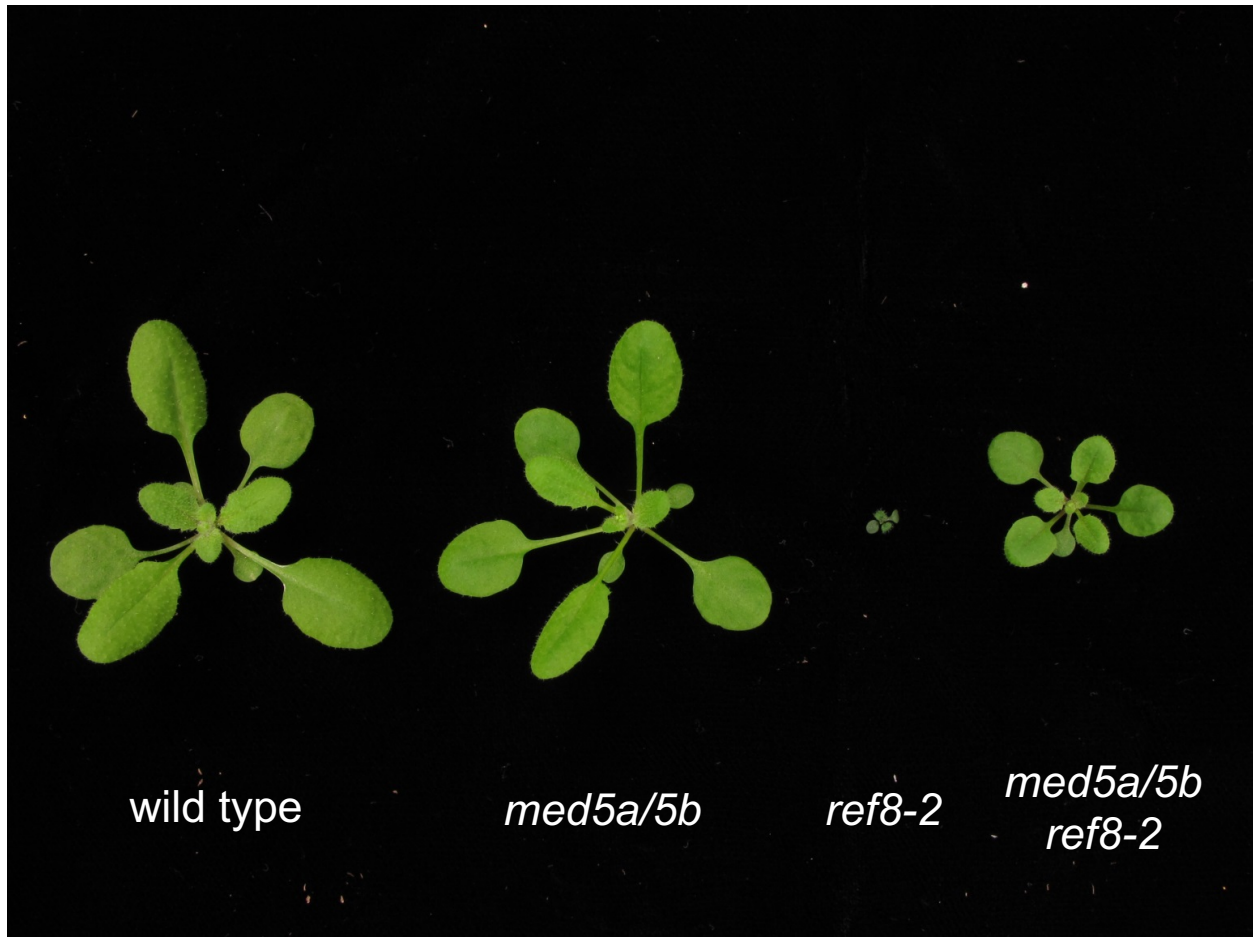


Figure 1.4 Knocking-out *MED5a* and *MED5b* substantially rescues the arrested growth of *Arabidopsis ref8-2* mutant

Disruption of *MED5a* and *MED5b* rescued *ref8-2* that contains a T-DNA insertion in the gene encoding C3'H. Wild-type, *med5a/5b*, *ref8-2*, and *med5a/5b ref8-2* plants were photographed four weeks after planting. The *ref8-2* mutant is dwarf and sterile, but the triple mutant *med5a/5b ref8-2* has restored growth and is fertile. Photograph provided by Nicholas D. Bonawitz.

CHAPTER 2. A ^{13}C ISOTOPE LABELING METHOD FOR THE MEASUREMENT OF LIGNIN METABOLIC FLUX IN ARABIDOPSIS STEMS

2.1 Abstract

2.1.1 Background

Metabolic fluxes represent the functional phenotypes of biochemical pathways and are essential to reveal the distribution of precursors among metabolic networks. Although analysis of metabolic fluxes, facilitated by stable isotope labeling and mass spectrometry detection, has been applied in the studies of plant metabolism, we lack experimental measurements for carbon flux towards lignin, one of the most abundant polymers in nature.

2.1.2 Results

We developed a feeding strategy of excised Arabidopsis stems with ^{13}C labeled phenylalanine (Phe) for the analysis of lignin biosynthetic flux. We optimized the feeding methods and found the stems continued to grow and lignify. Consistent with lignification profiles along the stems, higher levels of phenylpropanoids and activities of lignin biosynthetic enzymes were detected in the base of the stem. In the feeding experiments, ^{13}C labeled Phe was quickly accumulated and used for the synthesis of phenylpropanoid intermediates and lignin. The intermediates displayed two different patterns of labeling kinetics during the feeding period. Analysis of lignin showed rapid incorporation of label into all three subunits in the polymers.

2.1.3 Conclusions

Our feeding results demonstrate the effectiveness of the stem feeding system and suggest a potential application for the investigations of other aspects in plant metabolism. The supply of exogenous Phe leading to a higher lignin deposition rate indicates the availability of Phe is a determining factor for lignification rates.

2.1.4 Keywords

stable isotope labeling, stem feeding, lignin, phenylpropanoids

2.2 Background

Metabolic fluxes quantify the activities of biochemical networks that functionally integrate metabolites, enzymes, and their interactions [1]. Mathematical analysis of metabolic fluxes is becoming increasingly attractive to assess the responses of metabolic pathways in plants to genetic or environmental perturbations [2]. Stable isotope labeling experiments coupled with mass spectrometry (MS) quantification have been used to measure the flux within metabolic networks in various plant systems and expand our knowledge of metabolism and the control of flux [3, 4]. For example, Szecowka *et al.* supplied $^{13}\text{CO}_2$ to intact *Arabidopsis* rosettes to estimate intracellular fluxes of photosynthesis and central carbon metabolism [5]. Developing seeds of *Arabidopsis*, oilseed rape, and *Brassica napus* were fed with ^{13}C labeled glucose or sucrose to elucidate fluxes in lipid synthesis and other primary metabolism [6-8]. Flux analysis has also been applied to unravel phenylpropanoid metabolism in plants. Matsuda *et al.* utilized deuterium labeled Phe to analyze metabolic flux towards defense-related products in potato tubers upon wounding stress or elicitor treatment [9-11]. Similarly, labeled Phe was provided to petunia flowers to address flux distribution through CoA-dependent and -independent pathways for benzenoid biosynthesis [3, 12]. A recent study with labeled glucose revealed that expression of the *Arabidopsis* transcription factor MYB12 induces primary and specialized metabolism to enhance flavonoid production in tomato fruits [13]. In contrast, there have been no measurements of metabolic flux into lignin, the most abundant carbon sink derived from phenylpropanoid metabolism.

Lignin is a heterogeneous aromatic polymer that constitutes approximately 20-30% of carbon fixed by land plants [14]. Deposited together with polysaccharides in the plant secondary cell wall, lignin provides mechanical strength and hydrophobicity for plants to stand upright and to transport water and nutrients through vascular structures [15]. Although important for plants, lignin is a major recalcitrance factor for forage digestibility, efficient paper-pulping, and biofuel production [16]. Alteration of lignin content and monomer composition in natural mutants and genetically engineered plants improves the efficiency of utilization of lignocellulosic biomass [17, 18], and has thereby motivated many studies of lignin biosynthesis and its regulation. Lignin in dicots is derived mainly from three monomers, *p*-coumaryl alcohol (H lignin), coniferyl alcohol (G lignin), and sinapyl alcohol (S lignin), all of which are synthesized from the phenylpropanoid pathway (Figure 2.1) [15, 19, 20]. Phenylpropanoid metabolism starts with deamination of phenylalanine (Phe) by Phe ammonia lyase (PAL), the ring of which is subsequently hydroxylated

by three cytochrome P450 enzymes [21] and *O*-methylated [22, 23], with some of these steps occurring at the level of CoA and shikimate esters of the hydroxycinnamic acids [24-26]. Ultimately, CoA thioesters are the substrates for reduction to their corresponding aldehydes and finally the alcohols commonly known as monolignols. Despite a comprehensive knowledge of the kinetics of most of the lignin biosynthetic enzymes and steady-state accumulation of metabolites involved, we lack systematic and quantitative measurements of the carbon flux into and within this branched metabolic pathway. Accurate quantification and prediction of lignin biosynthetic flux *in planta* will guide our understanding of controlling steps in the pathway and facilitate future genetic manipulations of lignification.

Reliable estimation of phenylpropanoid flux towards lignin and further analysis of flux regulation depend on precise quantification of phenylpropanoids and their isotopologues from appropriate plant tissue fed with isotope-labeled precursor. Arabidopsis stems deposit lignin at substantial levels and thus make a good system in which to analyze lignin biosynthetic flux not only in wild-type plants, but also in genetically modified plants with perturbations in Phe synthesis or lignin biosynthesis itself [19, 27]. Plants transport water and nutrients via the transpiration stream, which continues to function after plant organs have been removed from the parent plant. This property permitted experiments in which isotope-labeled amino acids or NH_4NO_3 were fed to white lupin, wheat, poplar, and Arabidopsis to study nitrogen transport and metabolism [28-31]. Similarly, stable isotope feeding assays of excised Arabidopsis leaves or whole rosettes were developed for the investigation of carbon partitioning in sink and source leaves [32]. In this study, we fed excised inflorescence stems of Arabidopsis with $^{13}\text{C}_6$ -ring labeled Phe ($^{13}\text{C}_6$ -Phe) to trace phenylpropanoid metabolism. Our recently developed liquid chromatography coupled with tandem MS (LC-MS/MS) based phenylpropanoid metabolic profiling method provides fast and comprehensive analysis of pathway intermediates and end products [33]. All six isotopically-labeled carbons (^{13}C) in the Phe ring are maintained in all intermediates and in lignin, allowing downstream labeled compounds to be distinctly detected and accurately quantified by MS. Furthermore, incorporation of label into lignin can be measured to estimate flux to the major pathway end product. We have optimized the feeding conditions and analyzed metabolite pool sizes and isotopic abundances, selected enzyme activities and lignin deposition during the feeding process. Over a time course of only 360 min, $^{13}\text{C}_6$ -Phe rapidly labeled soluble intermediates and lignin, demonstrating the efficiency of the feeding strategy. These data together can be used for

mathematical modeling of lignin deposition and the exploration of the regulation of lignin biosynthesis in plants [3, 9-11, 12].

2.3 Results

2.3.1 Establishment of a stem feeding system using wild-type Arabidopsis

To perform stable isotope labeling experiments, we first established an experimental system using Arabidopsis stems that can be fed with exogenous [$^{13}\text{C}_6$]-Phe to synthesize labeled phenylpropanoids. We cut Arabidopsis stems under water and then transferred them into liquid Murashige and Skoog medium (MS medium) supplemented with [$^{13}\text{C}_6$]-Phe in 1.5 mL tubes (Figure 2.2, see Method for details.). Excising stems under water prevented cavitation of the xylem so that the transpiration stream would not be blocked.

To determine whether the excised stems are metabolically active we first asked whether the stems continue to grow in MS medium. When we assessed the growth of four-week-old inflorescence stems, they elongated 2 cm in 48 hours (Figure 2.3 A), the same as the reported growth rate of Arabidopsis stems in soil [34]. To test if the liquid medium is optimal for growth, stems were also grown in a series of diluted MS media. In the first 48 hours after excision, stems grew in all conditions and showed the largest elongation in full-strength MS medium, suggesting that the medium is adequate and not toxic to the excised stems (Figure 2.3 A). In order to accurately estimate metabolic fluxes in stems, it is important that the supplied Phe does not perturb their growth [1], so we next analyzed the elongation of stems incubated in MS medium containing Phe of different concentrations. Supplying 0.1 to 3 mM of Phe had no effect on stem growth over 48 hours (Figure 2.3 B). The amount of medium taken up by the stems was determined by weighing the medium remaining in the tube at various time points during the feeding process (Figure 2.4). On average, each excised stem took up $1.41 \pm 0.49 \text{ mg min}^{-1}$, equivalent to $1.4 \text{ } \mu\text{L min}^{-1}$ given that the density of medium is 1.0 g mL^{-1} . Loss of medium through evaporation was negligible. These data indicate that we can feed Arabidopsis stems with exogenous Phe in conventional MS medium.

We next addressed whether the excised Arabidopsis stems can utilize the supplied [$^{13}\text{C}_6$]-Phe to synthesize labeled lignin. We fed stems with 0.3 mM [$^{13}\text{C}_6$]-Phe for 48 hours and analyzed lignin monomers using the derivatization followed by reductive cleavage (DFRC) method coupled with gas chromatography (GC)/MS [35]. The DFRC method cleaves β -O-4 linkages within lignin

and can be used to detect monolignols incorporated into the polymer derived from [$^{13}\text{C}_6$]-Phe. Monomers released corresponding to H, G, and S subunits from the stems fed with [$^{13}\text{C}_6$]-Phe were labeled to 77%, 27%, and 22%, respectively (Figure 2.5). The higher labeling of H monomers released by DFRC relative to G and S may be due in part to the stems having larger pre-existing deposits of G and S lignin [19, 36], resulting in dilution of labeled monomers. The detection of $^{13}\text{C}_6$ -ring labeled lignin monomers in the stems indicates that the exogenous Phe is used for lignification and that $^{13}\text{C}_6$ -ring labeled lignin precursors are detectable by LC-MS/MS.

2.3.2 Phenylpropanoid metabolism is not homogeneous along the stem

The Arabidopsis inflorescence stem is a heterogeneous organ, in which the metabolite levels and enzyme activities have been reported to vary along the stem developmental gradient [27, 36]. For reliable flux measurements, an experimental tissue with consistent cellular phenylpropanoid concentrations and enzyme activities is needed. To determine the extent of heterogeneity of the Arabidopsis primary stem and find the best tissue for future experiments, we analyzed the concentrations of phenylpropanoids and their labeling percentage along the stems after feeding 0.3 mM of [$^{13}\text{C}_6$]-Phe for 360 min. The levels of endogenous compounds are similar along the stems (Figure 2.6 A), whereas, the concentrations of labeled compounds were much higher near the base of the stem than in the top. Over the six-hour period [$^{13}\text{C}_6$]-Phe was accumulated to 72 nmol g fresh weight (FW) $^{-1}$ in the basal 2 cm, but only reached 4 nmol g FW $^{-1}$ above 2 cm. Similarly, the concentration of [$^{13}\text{C}_6$]-*p*-coumarate was 12 nmol g FW $^{-1}$ at the base, but only 0.4 nmol g FW $^{-1}$ nearer the top. The high concentration of downstream intermediates including *p*-coumarate suggests that in the basal stem, exogenous Phe is actively sequestered from the transpiration stream, and possibly that enzymes in this region are more active in the consumption of Phe for lignin production than in upper parts. These data indicate that the basal 2 cm stem fragment appears to be an appropriate tissue in which to measure phenylpropanoid metabolic flux.

In addition to metabolite analysis, PAL and 4-coumarate: CoA ligase (4CL) enzyme activities were measured to assess the heterogeneity of lignin biosynthetic enzymes along the Arabidopsis inflorescence stem axis (Figure 2.6 B). PAL and 4CL both showed higher activities in the basal stems compared to those in the apical region. The enzyme activities together with the soluble metabolite pool sizes suggest that the stems have a larger phenylpropanoid flux in the base

than the top, again consistent with the stem base being the best tissue for analysis. We thus chose to harvest this basal 2 cm fragment of stems for analysis in our feeding experiments.

As shown above, phenylpropanoid metabolism varies across the developmental profile of the *Arabidopsis* inflorescence stem, but it is also known that metabolite accumulation, enzyme activities and gene expression can exhibit diurnal fluctuations. Recent studies have shown that transcript abundance of lignin biosynthetic genes is higher during the night in *Arabidopsis* [37] so we determined PAL and 4CL enzyme activities over a 24-hour period. We found that both enzymes were not statistically different in the basal stem 0-2 cm fragments (Figure 2.7). However, 4CL activity was higher at night compared to that in the day in the 2-4 cm fragment of the stem. These results indicate that it is appropriate to conduct our feeding experiment during the day and on 0-2 cm fragments.

2.3.3 Soluble phenylpropanoids are rapidly labeled in stems fed with [$^{13}\text{C}_6$]-Phe

Using the established system, we fed four-week-old wild-type *Arabidopsis* stems with 1.0 mM [$^{13}\text{C}_6$]-Phe and sampled basal 0-2 cm fragments over a time course of 360 min. [$^{13}\text{C}_6$]-Phe was used at a concentration of 1 mM to achieve high accumulation of labeled downstream phenylpropanoids for accurate quantification. Phe and the soluble phenylpropanoid intermediates were quantified by LC-MS/MS [33]. Supplied Phe was rapidly taken up by the stems as 33% of the entire Phe pool was labeled after 2 min (Figure 2.8). At this time point downstream products including *p*-coumarate and ferulate were also labeled. After 40 min, caffeate, *p*-coumaraldehyde, coniferaldehyde, sinapaldehyde, and the corresponding monolignols were all labeled indicating that exogenous [$^{13}\text{C}_6$]-Phe was rapidly transported into lignifying cells for phenylpropanoid metabolism.

During the labeling period of 360 min, Phe and intermediates in the early part of the pathway rapidly approached isotopic steady state, whereas downstream metabolites showed a slower increase in labeling enrichment (Figure 2.8). To achieve a quantitative overview of the labeling kinetics, we used hierarchical clustering to analyze the isotope enrichment profiles of the compounds measured. The hierarchical clustering based on squared Euclidian distance grouped the labeling percentage of phenylpropanoids into two patterns (Figure 2.9). The first group contained Phe, *p*-coumarate, caffeate, ferulate, *p*-coumaraldehyde, *p*-coumaryl alcohol, and *p*-coumaroyl shikimate each of which exhibited rapid isotope enrichment in the first 80 min and a

high final labeling of approximately 90%, except for *p*-coumaryl alcohol, which was 74%. In contrast, the second group including caffeoyl shikimate, coniferaldehyde, sinapaldehyde, coniferyl alcohol, sinapyl alcohol, and sinapate all showed slower increases in labeling, reaching less than 60% final labeling except for caffeoyl shikimate, which was slightly higher. These results are consistent with precursor-product relationships [38], where compounds closer to the fed substrate [$^{13}\text{C}_6$]-Phe are labeled faster and to a higher percentage than downstream product or than products with large endogenous pools. Furthermore, the efficient labeling of *p*-coumaryl alcohol provides an alternative or additional explanation for the high labeling of H-lignin that we previously observed (Figure 2.5).

In contrast to the pattern of isotopic enrichment observed for early pathway intermediates, the accumulation of labeled Phe and, except for sinapate, all downstream metabolites constantly increased and did not reach a metabolic steady state over the feeding period. Endogenous (unlabeled) Phe increased from 33 nmol g FW⁻¹ to around 50 nmol g FW⁻¹ after 80 min, while [$^{13}\text{C}_6$]-Phe rapidly accumulated and reached approximately 400 nmol g FW⁻¹ after 360 min, resulting in a labeling percentage of 89% (Figure 2.10). Similarly, endogenous *p*-coumarate and ferulate concentrations significantly increased over the time course. [$^{13}\text{C}_6$]-*p*-coumarate and [$^{13}\text{C}_6$]-ferulate accumulated to 90 and 10.7 nmol g FW⁻¹ after 360 min, approximately 20 times their endogenous levels (Figure 2.10). Neither endogenous cinnamate nor [$^{13}\text{C}_6$]-cinnamate were detected in the stem tissue, even when high concentrations of [$^{13}\text{C}_6$]-Phe and [$^{13}\text{C}_6$]-*p*-coumarate were observed, possibly because cinnamate 4-hydroxylase (C4H) efficiently catalyzes cinnamate hydroxylation such that the level of its substrate remains below the limits of detection. Endogenous caffeate was at or near the detection limits in our experiment, however [$^{13}\text{C}_6$]-caffeate could be quantified readily after 40 min of feeding and reached 1.5 nmol g FW⁻¹ after 360 min (Figure 2.10). Unlike the other hydroxycinnamic acids, endogenous sinapate concentration was not changed during the feeding process, and [$^{13}\text{C}_6$]-sinapate reached steady state after 80 min, at a level lower than the unlabeled sinapate (Figure 2.10). The pool size of endogenous unlabeled *p*-coumaraldehyde showed about a 3-fold increase after 360 min, while unlabeled *p*-coumaryl alcohol's concentration did not change. Similarly, unlabeled coniferyl alcohol, sinapyl alcohol and their corresponding aldehydes did not change in abundance. The accumulation of these labeled aldehydes and alcohols were relatively modest (Figure 2.10).

To test if the activities of biosynthetic enzymes are affected by the elevated phenylpropanoid levels, we determined the activities of PAL and 4CL in the stem tissue fed with 1 mM Phe over the time course. As shown in Figure 2.12, PAL and 4CL activities both stayed constant even when the feeding period was extended to 480 min. This result suggests that increased levels of phenylpropanoids did not induce PAL or 4CL activities during the feeding process.

2.3.4 Isotopic label is incorporated into lignin in stems fed with [$^{13}\text{C}_6$]-Phe

To examine the isotopic labeling of lignin in the feeding experiment, we measured lignin composition by DFRC/GC using flame ionization detection (FID), and the labeling percentage of each released monomer by DFRC/GC/MS in stems fed with 1 mM [$^{13}\text{C}_6$]-Phe. G and S lignin released monomers were labeled in the primary stems after 60 min, while labeled H lignin was observed after 120 min (Figure 2.13 A). In samples where labeled H lignin was detected, its labeling percentage was more than two-fold that of G or S lignin. This result is consistent with the observation that H lignin was labeled to a much higher level in whole stems fed for 48 hours (Figure 2.5). Next, we wanted to use the isotopic labeling of DFRC products to calculate the rate of lignin deposition during the feeding experiment. Because DFRC only detects lignin subunits that are exclusively β -O-4 linked (Fu et al., 1998), we measured lignin content by both the DFRC and the acetyl bromide methods and found that total lignin analyzed by the latter was 21.3-fold ($n = 15$) of DFRC lignin. We then used this value as a conversion factor to estimate lignin deposition using the DFRC data alone. A closer analysis of labeled H lignin content showed a rapid increase from near 80 nmol g FW $^{-1}$ at 120 min to over 1200 nmol g FW $^{-1}$ at 360 min (Figure 2.13 B). This large accumulation contributed to the high labeling percentage of 60% in addition to the small amount of endogenous unlabeled H lignin (Figure 2.13 A). Labeled G, and S lignin was deposited in a linear fashion over the time course at rates of approximately 6 nmol g FW $^{-1}$ min $^{-1}$ and 1 nmol g FW $^{-1}$ min $^{-1}$, respectively. It is interesting that the ratio of deposition rates of labeled G and S lignin (6.1:1) differed from the G/S ratio of unlabeled lignin (3.6:1) measured at 0 min (Figure 2.13 C). The disproportional synthesis of labeled lignin subunits suggested that flux distribution towards different branches was affected by increased supply of the common precursor Phe. The total labeled lignin constituted by all three monomers showed a synthesis rate of 9.8 nmol g FW $^{-1}$ min $^{-1}$ in the sampled stem fragments. Because the changes in the sum of pre-existing lignins and the newly synthesized lignins deposited during the 360 min feeding experiment were within the

standard deviation range of the DFRC results, the analysis of unlabeled lignin content showed no changes over the time course (Figure 2.13 C).

2.4 Discussion

We developed a stable isotope feeding system to measure phenylpropanoid flux in intact *Arabidopsis* stems. In comparison to previous tracer experiments for lignin biosynthesis using cell suspension cultures [39, 40], the whole stems in our system continue their normal growth and development (Figure 2.3), and conduct *in situ* phenylpropanoid metabolism and lignification (Figure 2.5). The excised wild-type stems, resembling the plants growing in soil, consume supplied [$^{13}\text{C}_6$]-Phe that rapidly and efficiently labels the soluble phenylpropanoid intermediates and lignin over a time course of 360 min. The measurement of metabolites together with enzyme activities for lignin biosynthesis will provide valuable *in planta* input for analysis of flux through the phenylpropanoid pathway.

Application of isotope labeling has a long history in investigation of phenylpropanoid metabolism. Half a century ago, ^{14}C labeled precursors were fed to various plants and cell suspension cultures to solve the structure of the pathway [39, 40]. Recently, the widespread use of MS technology has made it feasible to employ stable isotope labeling for the systematic estimation of metabolic fluxes in biochemical networks [3]. To analyze the flux through the branched lignin biosynthetic pathway in growing plants, an optimal feeding system needs to maintain physiologically normal conditions and active lignification. To this end, we chose intact *Arabidopsis* stems, where lignin is primarily deposited, excised from plants grown in soil. Previously, xylem sap extracted from plants was used as liquid medium to feed isotope labeled compounds to shoots of various plants [30, 31]. Because xylem sap contains amino acids including Phe [31, 41], we utilized a defined MS medium that avoids interference with endogenous unlabeled Phe and is practical to prepare in large volumes. The medium is adequate for stems to grow (Figure 2.3) and even to develop fertile seeds when fresh medium was provided to the stems every day (data not shown). As the cauline leaves were all left intact, their transpiration promoted the rapid absorption of liquid MS medium together with the [$^{13}\text{C}_6$]-Phe at a rate of $1.4\ \mu\text{L min}^{-1}$ (Figure 2.4). When fed with 1 mM [$^{13}\text{C}_6$]-Phe, each stem on average took up approximately 500 nmol labeled Phe after 360 min, equivalent to 10% of the carbon fixed by photosynthesis in an *Arabidopsis* rosette of 200 mg in the same amount of time [5]. At the end of the feeding experiment, the stem

base (approximately 30 mg) deposited over 3.5 $\mu\text{mol g FW}^{-1}$ labeled lignin (Figure 2.13), accounting for more than 20% of the labeled Phe absorbed. This measurement of the end products gives an estimation of total flux entering the pathway.

The Arabidopsis stem is a complex organ that comprises different tissues and cells, each of which differs in the level of phenylpropanoids they synthesize and the nature of the end products they accumulate. To find a suitable tissue for analysis, we assessed phenylpropanoid metabolite content and enzyme activities along the inflorescence stems and throughout the feeding process. In these experiments we found that the base of the stems contained the highest levels of labeled precursor and activities of PAL and 4CL (Figure 2.6), demonstrating a high phenylpropanoid flux, consistent with the lignin deposition pattern observed in Arabidopsis [42].

The feeding of [$^{13}\text{C}_6$]-Phe to Arabidopsis stems showed rapid incorporation of label into downstream phenylpropanoid metabolism. Like previous studies done in various plant systems [43-46], we observed quick and high accumulation of labeled hydroxycinnamic acids (Figure 2.10). The labeling kinetics of *p*-coumarate and ferulate followed that of Phe, indicative of rapid isotope incorporation. Interestingly, the labeling percentage profiles of all phenylpropanoids measured showed two different patterns apparently based on whether or not their synthesis was dependent on the enzyme *p*-coumaroylshikimate 3'-hydroxylase (C3'H). The compounds that do not require C3'H for their synthesis displayed rapid increases in labeling enrichment and reach isotopic steady state within 80 min of feeding. In contrast, C3'H-dependent compounds were mostly labeled more slowly (Figure 2.8). One possible explanation is that the relatively large pools of endogenous unlabeled coniferyl alcohol, sinapyl alcohol, and sinapate in the stems render the labeling slow, but it cannot explain the results seen with coniferaldehyde and sinapaldehyde which had much less accumulation (Figure 2.10). Another possibility is that a slow (often rate-limiting) enzymatic step provokes high labeling of upstream intermediates and causes slow and low labeling in downstream products [4, 47]. C3'H is a cytochrome P450-dependent monooxygenase, a class of enzymes generally known to have low turnover [48, 49]. In the feeding experiments, the substantial accumulation of *p*-coumaroylshikimate indicates that C3'H might be saturated, becoming a limiting step for lignification. Caffeate and ferulate are two exceptions in the first group as both are 3'-hydroxylated but exhibited a labeling pattern like Phe and *p*-coumarate. The labeling percentage of ferulate is higher than its shikimate ester precursors, which seems to suggest *p*-coumarate is a close upstream precursor for caffeate and ferulate and that its conversion to these

two downstream intermediates bypasses the HCT/C3'H/HCT/CSE steps. Supporting this model, it was shown that C4H and C3'H from poplar form a protein complex and that the recombinant proteins can catalyze the hydroxylation of *p*-coumarate *in vitro* [50].

Lignification in plants is tightly regulated transcriptionally and post-translationally to control carbon allocation into this major metabolic sink. Recent studies have shown that the expression of *PAL* genes is regulated by Mediator and transcriptional factors [51, 52] and that Kelch repeat F-box proteins mediate the degradation of PAL enzymes via the proteasome complex [53]. PAL has often been considered as the rate-limiting step in lignin biosynthesis [54] and this may well be the case when PAL is down-regulated. In contrast, our study shows that in wild-type *Arabidopsis* plants, PAL is present in excess and the availability of Phe instead determines lignin deposition, a factor that has been undervalued in the past. Before feeding, stems had developed for 6 days in soil and deposited approximately 12 $\mu\text{mol g FW}^{-1}$ of lignin. When fed with 1.0 mM [$^{13}\text{C}_6$]-Phe for only 360 min, stems synthesized over 3 $\mu\text{mol g FW}^{-1}$ additional, labeled, lignin, accounting for 25% of the previous total, showing a higher lignification rate than would occur under normal, unfed, conditions. Given that PAL activity remains unaltered during the feeding period (Figure 2.12), these data indicate that an enhanced supply of Phe increases lignin biosynthesis in plants, and that PAL is not saturated with endogenous Phe *in vivo* under normal conditions. Consistent with these observations, it has been reported that the *Arabidopsis pall* mutant has a 40% reduction in PAL activity but deposits wild-type levels of lignin [55, 56], supporting the idea that PAL in wild-type plants does not perform at maximal velocity and is not the rate-limiting enzyme.

Many studies have reported that lignin monomer composition is altered in transgenic plants or mutants when flux into or within the phenylpropanoid pathway is reduced. In addition to decreased lignin content, an increased S/G ratio was seen in mutants with reduced levels of PAL, C4H, 4CL, or aroclate dehydratase, an enzyme in Phe biosynthesis [55, 57-59]. It has been suggested that these changes occur when enzymes with markedly different kinetic characteristics that act at pathway branch points compete for a common intermediate, thus leading to altered flux partitioning when the two enzymes' substrate concentration is decreased [58]. In contrast, experimental evidence to show the opposite change, increased total influx leading to a decreased S/G ratio, has not previously been reported. As this model predicts, we found that enhanced supply of Phe resulted in increased H lignin deposition and comparatively reduced S lignin synthesis

(Figure 2.13). In plants, cinnamoyl CoA reductase (CCR) competes with hydroxycinnamoyl CoA: shikimate hydroxycinnamoyl transferase (HCT) for *p*-coumaroyl CoA to drive flux towards H lignin (Figure 2.1). It has been suggested that HCT associates with the complex of C4H and C3'H on the endoplasmic reticulum membrane, forming a metabolon that efficiently directs carbon towards G and S lignin in plants under normal conditions [60]. This is suggested to occur even though the reported K_M of Arabidopsis CCR1 (2.27 μ M towards *p*-coumaroyl CoA, [61]) is smaller than the K_M of HCT from tobacco (600 μ M, [23]) and the maximal activity of CCR (200 pkat mg⁻¹ protein) in Arabidopsis stem extracts is higher than that of HCT (35 pkat mg⁻¹ protein) using *p*-coumaroyl CoA as substrate (unpublished data). When *p*-coumaroyl CoA accumulates and C3'H in the metabolon becomes saturated, CCR likely outcompetes HCT and generates higher levels of *p*-coumaraldehyde, which ultimately results in more H lignin. Similarly, ferulate 5-hydroxylase, the last cytochrome P450-dependent monooxygenase in the pathway, typically competes effectively with cinnamyl alcohol dehydrogenase and the yet-to-be identified transporter of coniferyl alcohol to synthesize S lignin (Figure 2.1) [62]. When high concentrations of coniferaldehyde and coniferyl alcohol are present, proportionally more flux escapes 5-hydroxylation leading to a decreased S/G ratio (Figure 2.13). The precise control of flux by the enzymes at each branch points will be better understood by a more sophisticated and mechanistic mathematic model that includes the kinetics of the enzymes involved.

2.5 Conclusions

In summary, we established an experimental feeding system to supply intact Arabidopsis stems with [¹³C₆]-Phe to investigate the metabolic flux towards lignin. Our analysis revealed that the availability of Phe determines lignin deposition rate and can alter distribution of flux towards three monolignols. The soluble phenylpropanoid metabolite and lignin measurements from dynamic isotope labeling experiments can be input for mathematical modeling of metabolic fluxes to quantitatively unravel the control of flux and to explore its regulation. In addition to the measurement of lignin biosynthetic flux in wild-type Arabidopsis, we envision a wider application of this stem feeding system. For example, mutants with defects in lignification can be fed with labeled precursor in parallel with the wild type to explore how genetic perturbation affects flux distribution within the phenylpropanoid pathway. Furthermore, given the abundance of Arabidopsis mutants in myriad biochemical pathways, isotope labeled substrates could be

administered with this system such that their flux distribution could be described but in this case, reoptimization of the selected tissue would be warranted.

2.6 Methods

2.6.1 Plant material and growth conditions

Arabidopsis thaliana (Columbia-0) were grown in soil at 23 °C under light intensity of 150 $\mu\text{E m}^{-2} \text{s}^{-1}$ in a photoperiod of 16-hour light and 8-hour dark in a growth chamber. Primary stems of four-week-old plants were used for feeding experiments and enzyme activity assays.

2.6.2 Feeding

Arabidopsis stems were fed with liquid MS medium containing [$^{13}\text{C}_6$]-Phe in 1.5 mL tubes in the same growth chamber as they were grown in to maintain the same environmental conditions. [$^{13}\text{C}_6$]-Phe was filter-sterilized before being added into autoclaved MS medium. A steel nail was heated and used to form a hole in the lid of tubes so that the stems could be inserted through it. Plants were next removed from soil, and then the inflorescence stem was cut immediately above the rosette with a double-edged razor blade under water. The excised stem was immediately transferred into incubation solution and placed on a rack in the growth chamber (Figure 2.2). Tubes were preloaded with 0.5 mL medium and refilled with fresh medium when necessary. After the specified incubation period, stem fragments were rinsed three times with water to remove [$^{13}\text{C}_6$]-Phe on the surface and frozen in liquid nitrogen. Ten stem fragments were harvested as one sample, and three biological replicates were generated for each time point. Depending on the biological questions asked and quantification methods employed, more biological replicates or stem tissue *per* sample may be optimized to achieve accurate measurements of metabolites. Samples were stored at -70 °C until extraction.

2.6.3 Stem growth measurement

Stems of four-week-old plants were cut under water as described above and used for growth analysis. Stems were photographed alongside a ruler immediately and after 3, 6, 9, 12, 24, 27, 30, 33, 36, and 48 hours. Measurements of stem length from the base to the apical meristem were done using ImageJ in comparison to the ruler included in the photograph.

2.6.4 LC-MS/MS analysis of soluble metabolites

Soluble metabolites were analyzed with LC-MS/MS following the method developed by Jaini *et al.* [33]. Briefly, stem tissue was extracted in 75% methanol (v/v) (at 10 $\mu\text{L mg FW}^{-1}$) for 2 hours at 65 °C, and supernatant was collected after centrifugation for 20 min at 16,000 x g. To concentrate the samples, 500 μL of supernatant was dried in speed-vac and then re-dissolved in 50 μL 50% methanol (v/v). Samples were analyzed by LC-MS/MS using a Zorbax Eclipse C8 column (150 mm 4.6 mm, 5 μm , Agilent Technologies, Santa Clara, CA) and ammonium acetate (pH 5.6) and acetonitrile/ H_2O / HCOOH (9.8/2/0.2) as mobile phase. Metabolite detection was achieved with a QTrap 5500 triple quadrupole mass spectrometer (AB Sciex, Redwood City, CA), equipped with an ESI-Turbolon-spray to operate in negative ion mode. Multiple reaction monitoring mode was used to quantify compounds. Standards of Phe, cinnamate, *p*-coumarate, caffeate, ferulate, sinapate, shikimate, *p*-coumaraldehyde, coniferaldehyde, sinapaldehyde, *p*-coumaryl alcohol, coniferyl alcohol, sinapyl alcohol, *p*-coumaroyl shikimate, caffeoyl shikimate were analyzed by LC-MS/MS to generate calibration curve to quantify the soluble metabolites. The same curve was used to quantify both unlabeled and labeled isotopologues. The labeling percentage was calculated with the following equation:

$$\text{labeling percentage} = \frac{[\text{labeled compound}]}{[\text{labeled compound}] + [\text{unlabeled compound}]} * 100\%$$

The hierarchical clustering analysis of isotope enrichment profiles of all measured metabolites was performed using averaged labeling percentage data of all time points. Squared Euclidian distance was computed and clustered in R [63].

2.6.5 DFRC/GC/FID and DFRC/GC/MS

Stems were analyzed for lignin content following the method established by Lu and Ralph, 1997 [35]. Briefly, samples were ground in liquid nitrogen, washed five times with 10 mL 70% ethanol (v/v) at 80 °C and once with acetone. The dried cell wall residue was weighed and dissolved in 2.5 mL of acetyl bromide/glacial acetic acid (20:80, v/v) mixture containing 0.2 mg internal standard (4, 4'-ethyldenebisphenol) overnight at room temperature. The mixture was dried under nitrogen gas then dissolved in 2 mL of dioxane/glacial acetic acid/ H_2O (50:40:10, v/v/v). 50 mg zinc dust was added to the mixture, vortexed, stirred for 25 min and applied to a solid phase extraction column (Discovery® DSC-18 SPE tube) pre-conditioned with 95% ethanol (v/v) and

H₂O. The column was washed with 5 mL 25% ethanol (v/v) and DFRC reactions products were eluted with 2.5 mL 96% ethanol (v/v) then dried under nitrogen gas. The sample was acetylated with 0.5 mL acetic anhydride/pyridine (60/40, v/v) overnight and dried again under nitrogen gas. The sample was then dissolved in 200 μ L dichloromethane and 1 μ L of the final product analyzed by GC/FID or GC/MS.

2.6.6 Acetyl bromide lignin analysis

Total lignin content in the stem tissue was quantified using acetyl bromide lignin analysis method described in [64] with minor revisions. Briefly, the basal 0-2 cm fragments of stems of four-week-old *Arabidopsis* were harvested and 15 biological replicates were included for analysis. The stem tissue was ground in liquid nitrogen and washed in 0.1 M sodium phosphate buffer (pH 7.2) at 50 °C for 30 min, followed by five washes with 10 mL 70% ethanol (v/v) at 80 °C and once with 2 mL acetone. The dried cell wall residue was weighed and dissolved in 2.5 mL of acetyl bromide/glacial acetic acid (25:75, v/v) overnight at room temperature. A control without cell wall residue was included. The dissolved samples were completely transferred into 50 mL volumetric flasks containing 2.5 mL 2 M NaOH and 12 mL acetic acid. 0.5 mL freshly prepared 7.5 M hydroxylamine hydrochloride was added into each sample followed by 35 mL acetic acid. The sample was mixed and allowed to settle before the volume was brought to 50 mL with acetic acid. Absorbance at 280 nm was measured on spectrophotometer using the control as blank. Extinction coefficient 23.20 g⁻¹ L cm⁻¹ was used to calculate total lignin content.

2.6.7 PAL and 4CL enzyme assays

Stem tissue was harvested and frozen in liquid nitrogen. Crude protein was extracted from ground tissue with Tris-HCl buffer at pH of 7.8, and desalted on a gel filtration column (SephadexTM, G-50 fine, GE Healthcare). PAL and 4CL assays were conducted following the method in Klempien 2010 [65]. Each PAL assay contained 100 mM Tris-HCl buffer pH 7.5, 5 mM Phe, and 5 μ L protein extract in a final volume of 50 μ L. The reactions were incubated at 23 °C for 120 min. The 4CL assay contained 100 mM Tris-HCl buffer pH 7.5, 5 mM MgCl₂, 5 mM ATP, 1 mM *p*-coumarate, 0.3 mM CoA, and 2 μ L protein extract in a final volume of 40 μ L. Each reaction was incubated at 23 °C for 20 min. Assay products were quantified on HPLC with

cinnamate and synthesized *p*-coumaroyl CoA as standards, respectively. Protein concentrations were measured with Bradford assay using bovine serum albumin as standard.

2.7 References

1. Sauer U: Metabolic networks in motion: ^{13}C -based flux analysis. *Mol Syst Biol* 2006, 2:62.
2. Libourel IGL, Shachar-Hill Y: Metabolic flux analysis in plants: From intelligent design to rational engineering. *Annu Rev Plant Biol* 2008, 59:625-650.
3. Colón AJM, Morgan JA, Dudareva N, Rhodes D: Application of dynamic flux analysis in plant metabolic networks. In: Schwender J, editor. *Plant Metabolic Networks*. Springer, New York, NY 2009:285-305.
4. Roscher A, Kruger NJ, Ratcliffe RG: Strategies for metabolic flux analysis in plants using isotope labelling. *J Biotechnol* 2000, 77(1):81-102.
5. Szecowka M, Heise R, Tohge T, Nunes-Nesi A, Vosloh D, Huege J, Feil R, Lunn J, Nikoloski Z, Stitt M *et al*: Metabolic fluxes in an illuminated *Arabidopsis* rosette. *Plant Cell* 2013, 25(2):694-714.
6. Schwender J, Hebbelmann I, Heinzl N, Hildebrandt T, Rogers A, Naik D, Klapperstuck M, Braun HP, Schreiber F, Denolf P *et al*: Quantitative multilevel analysis of central metabolism in developing oilseeds of oilseed rape during in vitro culture. *Plant Physiol* 2015, 168(3):828-848.
7. Lonien J, Schwender J: Analysis of metabolic flux phenotypes for two *Arabidopsis* mutants with severe impairment in seed storage lipid synthesis. *Plant Physiol* 2009, 151(3):1617-1634.
8. Hay JO, Shi H, Heinzl N, Hebbelmann I, Rolletschek H, Schwender J: Integration of a constraint-based metabolic model of *Brassica napus* developing seeds with (^{13}C) -metabolic flux analysis. *Front Plant Sci* 2014, 5:724.
9. Heinzle E, Matsuda F, Miyagawa H, Wakasa K, Nishioka T: Estimation of metabolic fluxes, expression levels and metabolite dynamics of a secondary metabolic pathway in potato using label pulse-feeding experiments combined with kinetic network modelling and simulation. *Plant J* 2007, 50(1):176-187.

10. Matsuda F, Morino K, Ano R, Kuzawa M, Wakasa K, Miyagawa H: Metabolic flux analysis of the phenylpropanoid pathway in elicitor-treated potato tuber tissue. *Plant Cell Physiol* 2005, 46(3):454-466.
11. Matsuda F, Morino K, Miyashita M, Miyagawa H: Metabolic flux analysis of the phenylpropanoid pathway in wound-healing potato tuber tissue using stable isotope-labeled tracer and LC-MS spectroscopy. *Plant Cell Physiol* 2003, 44(9):961-961.
12. Boatright J, Negre F, Chen XL, Kish CM, Wood B, Peel G, Orlova I, Gang D, Rhodes D, Dudareva N: Understanding in vivo benzenoid metabolism in petunia petal tissue. *Plant Physiol* 2004, 135(4):1993-2011.
13. Zhang Y, Butelli E, Alseekh S, Tohge T, Rallapalli G, Luo J, Kwar PG, Hill L, Santino A, Fernie AR *et al*: Multi-level engineering facilitates the production of phenylpropanoid compounds in tomato. *Nat Commun* 2015, 6:8635.
14. Pauly M, Keegstra K: Cell-wall carbohydrates and their modification as a resource for biofuels. *Plant J* 2008, 54(4):559-568.
15. Boerjan W, Ralph J, Baucher M: Lignin biosynthesis. *Annu Rev Plant Biol* 2003, 54:519-546.
16. Li X, Weng JK, Chapple C: Improvement of biomass through lignin modification. *Plant J* 2008, 54(4):569-581.
17. Wang P, Dudareva N, Morgan JA, Chapple C: Genetic manipulation of lignocellulosic biomass for bioenergy. *Curr Opin Chem Biol* 2015, 29:32-39.
18. Chen F, Dixon RA: Lignin modification improves fermentable sugar yields for biofuel production. *Nat Biotechnol* 2007, 25(7):759-761.
19. Bonawitz ND, Chapple C: The genetics of lignin biosynthesis: connecting genotype to phenotype. *Annu Rev Genet* 2010, 44:337-363.
20. Humphreys JM, Chapple C: Rewriting the lignin roadmap. *Curr Opin Plant Biol* 2002, 5(3):224-229.
21. Bonawitz ND, Chapple C: Can genetic engineering of lignin deposition be accomplished without an unacceptable yield penalty? *Curr Opin Biotechnol* 2013, 24(2):336-343.
22. Vanholme R, Cesarino I, Rataj K, Xiao YG, Sundin L, Goeminne G, Kim H, Cross J, Morreel K, Araujo P *et al*: Caffeoyl shikimate esterase (CSE) is an enzyme in the lignin biosynthetic pathway in Arabidopsis. *Science* 2013, 341(6150):1103-1106.

23. Goujon T, Ferret V, Mila I, Pollet B, Ruel K, Burlat V, Joseleau JP, Barriere Y, Lapierre C, Jouanin L: Down-regulation of the *AtCCR1* gene in *Arabidopsis thaliana*: effects on phenotype, lignins and cell wall degradability. *Planta* 2003, 217(2):218-228.
24. Mir Derikvand M, Sierra JB, Ruel K, Pollet B, Do CT, Thevenin J, Buffard D, Jouanin L, Lapierre C: Redirection of the phenylpropanoid pathway to feruloyl malate in *Arabidopsis* mutants deficient for cinnamoyl-CoA reductase 1. *Planta* 2008, 227(5):943-956.
25. Franke R, Hemm MR, Denault JW, Ruegger MO, Humphreys JM, Chapple C: Changes in secondary metabolism and deposition of an unusual lignin in the *ref8* mutant of *Arabidopsis*. *Plant J* 2002, 30(1):47-59.
26. Hoffmann L, Maury S, Martz F, Geoffroy P, Legrand M: Purification, cloning, and properties of an acyltransferase controlling shikimate and quinate ester intermediates in phenylpropanoid metabolism. *J Biol Chem* 2003, 278(1):95-103.
27. Vanholme R, Storme V, Vanholme B, Sundin L, Christensen JH, Goeminne G, Halpin C, Rohde A, Morreel K, Boerjan W: A systems biology view of responses to lignin biosynthesis perturbations in *Arabidopsis*. *Plant Cell* 2012, 24(9):3506-3529.
28. Sharkey PJ, Pate JS: Selectivity in xylem to phloem transfer of amino-acids in fruiting shoots of white lupin (*Lupinus albus* L.). *Planta* 1975, 127(3):251-262.
29. Vogelmann TC, Dickson RE, Larson PR: Comparative distribution and metabolism of xylem-borne amino-compounds and sucrose in shoots of *Populus deltoides*. *Plant Physiol* 1985, 77(2):418-428.
30. Simpson RJ, Lambers H, Dalling MJ: Nitrogen redistribution during grain growth in wheat (*Triticum aestivum* L.) .4. development of a quantitative model of the translocation of nitrogen to the grain. *Plant Physiol* 1983, 71(1):7-14.
31. Zhang LZ, Tan QM, Lee R, Trethewy A, Lee YH, Tegeder M: Altered xylem-phloem transfer of amino acids affects metabolism and leads to increased seed yield and oil content in *Arabidopsis*. *Plant Cell* 2010, 22(11):3603-3620.
32. Dethloff F, Orf I, Kopka J. Rapid in situ ¹³C tracing of sucrose utilization in *Arabidopsis* sink and source leaves. *Plant Methods* 2017 18:13-87.

33. Jaini R, Wang P, Dudareva N, Chapple C, Morgan JA: Targeted metabolomics of the phenylpropanoid pathway in *Arabidopsis thaliana* using reversed phase liquid chromatography coupled with tandem mass spectrometry. *Phytochem Anal* 2017, 28(4):267-276.
34. Suh MC, Samuels AL, Jetter R, Kunst L, Pollard M, Ohlrogge J, Beisson F: Cuticular lipid composition, surface structure, and gene expression in Arabidopsis stem epidermis. *Plant Physiol* 2005, 139(4):1649-1665.
35. Lu F, Ralph J: Derivatization followed by reductive cleavage (DFRC method), a new method for lignin analysis: protocol for analysis of DFRC monomers. *J Agric Food Chem* 1997, 45(7):2590-2592.
36. Meyer K, Shirley AM, Cusumano JC, Bell-Lelong DA, Chapple C: Lignin monomer composition is determined by the expression of a cytochrome P450-dependent monooxygenase in Arabidopsis. *Proc Natl Acad Sci USA* 1998, 95(12):6619-6623.
37. Mockler TC, Michael TP, Priest HD, Shen R, Sullivan CM, Givan SA, McEntee C, Kay SA, Chory J: The Diurnal project: diurnal and circadian expression profiling, model-based pattern matching, and promoter analysis. *Cold Spring Harb Symp Quant Biol* 2007, 72:353-363.
38. Ratcliffe RG, Shachar-Hill Y: Measuring multiple fluxes through plant metabolic networks. *Plant J* 2006, 45(4):490-511.
39. Brown SA: Lignins. *Annu Rev Plant Physiol* 1966, 17:223-244.
40. Hahlbrock K, Grisebach H: Enzymic controls in the biosynthesis of lignin and flavonoids. *Annu Rev Plant Physiol* 1979, 30:105-130.
41. Urquhart AA, Joy KW: Transport, metabolism, and redistribution of xylem-borne amino acids in developing pea shoots. *Plant Physiol* 1982, 69(5):1226-1232.
42. Rogers LA, Campbell MM: The genetic control of lignin deposition during plant growth and development. *New Phytol* 2004, 164(1):17-30.
43. Mccalla DR, Neish AC: Metabolism of phenylpropanoid compounds in *Salvia*. 2. biosynthesis of phenolic cinamic acids. *Can J Biochem Physiol* 1959, 37(4):537-547.
44. Chapple C, Vogt T, Ellis BE, Somerville CR: An Arabidopsis mutant defective in the general phenylpropanoid pathway. *Plant Cell* 1992, 4(11):1413-1424.

45. Higuchi T, Brown SA: Studies of lignin biosynthesis using isotopic carbon .12. biosynthesis and metabolism of sinapic acid. *Can J Biochem Physiol* 1963, 41(3):613-620.
46. Rasmussen S, Dixon RA: Transgene-mediated and elicitor-induced perturbation of metabolic channeling at the entry point into the phenylpropanoid pathway. *Plant Cell* 1999, 11(8):1537-1551.
47. Hanson AD, Rhodes D: ^{14}C tracer evidence for synthesis of choline and betaine via phosphoryl base intermediates in salinized sugarbeet leaves. *Plant Physiol* 1983, 71(3):692-700.
48. Schoch G, Goepfert S, Morant M, Hehn A, Meyer D, Ullmann P, Werck-Reichhart D: CYP98A3 from *Arabidopsis thaliana* is a 3'-hydroxylase of phenolic esters, a missing link in the phenylpropanoid pathway. *J Biol Chem* 2001, 276(39):36566-36574.
49. Mahesh V, Million-Rousseau R, Ullmann P, Chabrillange N, Bustamante J, Mondolot L, Morant M, Noirot M, Hamon S, de Kochko A *et al*: Functional characterization of two *p*-coumaroyl ester 3'-hydroxylase genes from coffee tree: evidence of a candidate for chlorogenic acid biosynthesis. *Plant Mol Biol* 2007, 64(1-2):145-159.
50. Chen HC, Li QZ, Shuford CM, Liu J, Muddiman DC, Sederoff RR, Chiang VL: Membrane protein complexes catalyze both 4-and 3-hydroxylation of cinnamic acid derivatives in monolignol biosynthesis. *Proc Natl Acad Sci USA* 2011, 108(52):21253-21258.
51. Bonawitz ND, Kim JI, Tobimatsu Y, Ciesielski PN, Anderson NA, Ximenes E, Maeda J, Ralph J, Donohoe BS, Ladisch M *et al*: Disruption of Mediator rescues the stunted growth of a lignin-deficient *Arabidopsis* mutant. *Nature* 2014, 509(7500):376-380.
52. Dolan WL, Dilkes BP, Stout JM, Bonawitz NB, Chapple C: Mediator complex subunits MED2, MED5, MED16, and MED23 genetically interact in the regulation of phenylpropanoid biosynthesis. *Plant Cell* 2017, 29(12):326903285.
53. Zhang XB, Gou MY, Guo CR, Yang HJ, Liu CJ: Down-regulation of kelch domain-containing f-box protein in *Arabidopsis* enhances the production of (poly)phenols and tolerance to ultraviolet radiation. *Plant Physiol* 2015, 167(2):337-548.
54. Bate NJ, Orr J, Ni WT, Meromi A, Nadlerhassar T, Doerner PW, Dixon RA, Lamb CJ, Elkind Y: Quantitative relationship between phenylalanine ammonia-lyase levels and phenylpropanoid accumulation in transgenic tobacco identifies a rate-determining step in natural product synthesis. *Proc Natl Acad Sci USA* 1994, 91(16):7608-7612.

55. Rohde A, Morreel K, Ralph J, Goeminne G, Hostyn V, De Rycke R, Kushnir S, Van Doorselaere J, Joseleau JP, Vuylsteke M *et al*: Molecular phenotyping of the *pal1* and *pal2* mutants of *Arabidopsis thaliana* reveals far-reaching consequences on phenylpropanoid, amino acid, and carbohydrate metabolism. *Plant Cell* 2004, 16(10):2749-2771.
56. Van Acker R, Vanholme R, Storme V, Mortimer JC, Dupree P, Boerjan W: Lignin biosynthesis perturbations affect secondary cell wall composition and saccharification yield in *Arabidopsis thaliana*. *Biotechnol Biofuels* 2013, 6.
57. Corea ORA, Ki C, Cardenas CL, Kim SJ, Brewer SE, Patten AM, Davin LB, Lewis NG: Arogenate dehydratase isoenzymes profoundly and differentially modulate carbon flux into lignins. *J Biol Chem* 2012, 287(14):11446-11459.
58. Li Y, Kim JI, Pysh L, Chapple C: Four isoforms of Arabidopsis 4-coumarate: CoA ligase have overlapping yet distinct roles in phenylpropanoid metabolism. *Plant Physiol* 2015, 169(4):2409-2421.
59. Schilmiller AL, Stout J, Weng JK, Humphreys J, Ruegger MO, Chapple C: Mutations in the cinnamate 4-hydroxylase gene impact metabolism, growth and development in *Arabidopsis*. *Plant J* 2009, 60(5):771-782.
60. Bassard JE, Richert L, Geerinck J, Renault H, Duval F, Ullmann P, Schmitt M, Meyer E, Mutterer J, Boerjan W, De Jaeger D, Mely Y, Goossens A, Werck-Reichhart D: Protein-protein and protein-membrane associations in the lignin pathway. *Plant Cell* 2012 24:4465–4482.
61. Baltas M, Lapeyre C, Bedos-Belval F, Maturano A, Saint-Aguet P, Roussel L, Duran H, Grima-Pettenati J: Kinetic and inhibition studies of cinnamoyl-CoA reductase 1 from *Arabidopsis thaliana*. *Plant Physiol Biochem* 2005, 43(8):746-753.
62. Humphreys JM, Hemm MR, Chapple C: New routes for lignin biosynthesis defined by biochemical characterization of recombinant ferulate 5-hydroxylase, a multifunctional cytochrome P450-dependent monooxygenase. *Proc Natl Acad Sci USA* 1999, 96(18):10045-10050.
63. R Core Team: R: a language and environment for statistical computing. <http://www.R-project.org> 2017.

64. Chang X, Chandra R, Berleth T, Beatson RP. Rapid, microscale, acetyl bromide-based method for high-throughput determination of lignin content in *Arabidopsis thaliana*. *J Agric Food Chem* 2005, 56 (16):6825-6834.
65. Klempien A, Kaminaga Y, Qualley A, Nagegowda DA, Widhalm JR, Orlova I, Shasany AK, Taguchi G, Kish CM, Cooper BR, D'Auria JC, Rhodes D, Pichersky E, Dudareva N. Contribution of CoA ligases to benzenoid biosynthesis in petunia flowers. *Plant Cell* 2012, 24(5):2015-30.

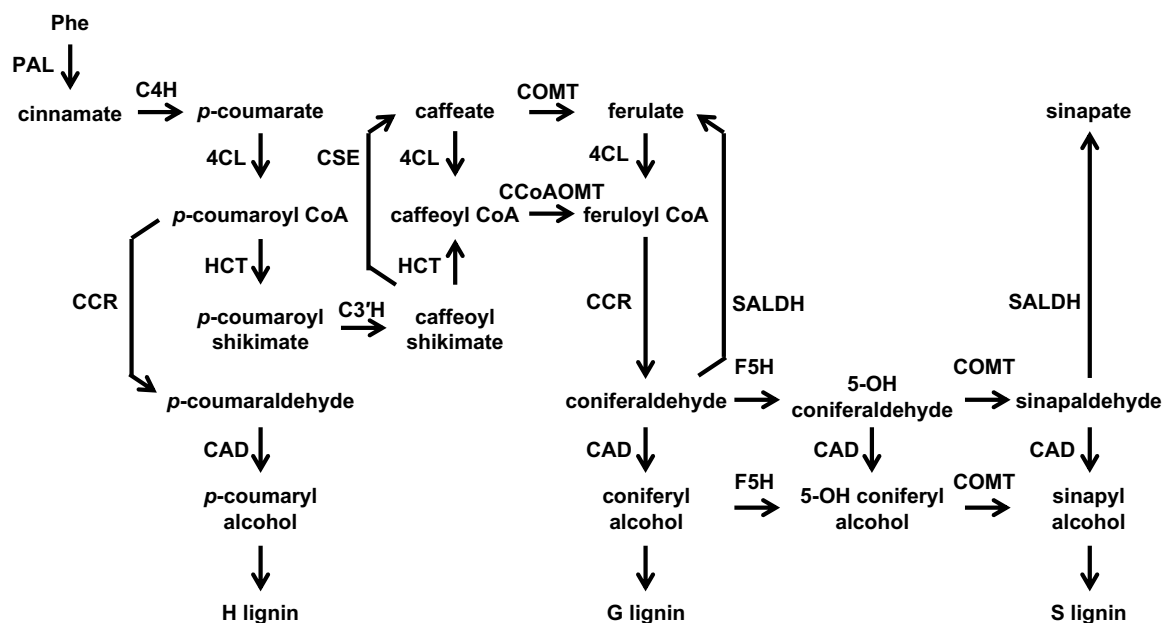


Figure 2.1 A simplified pathway illustrating lignin biosynthesis.

PAL, phenylalanine ammonia lyase; C4H, cinnamate 4-hydroxylase; 4CL, 4-coumarate CoA ligase; HCT, hydroxycinnamoyl CoA:shikimate hydroxycinnamoyl transferase; C3'H, *p*-coumaroyl shikimate 3'-hydroxylase; CSE, caffeoyl shikimate esterase; CCoAOMT, caffeoyl CoA *O*-methyltransferase; F5H, ferulate 5-hydroxylase; COMT, caffeic acid *O*-methyltransferase; CCR, cinnamoyl CoA reductase; CAD, cinnamyl alcohol dehydrogenase. SALDH, sinapaldehyde dehydrogenase.

A



B



C

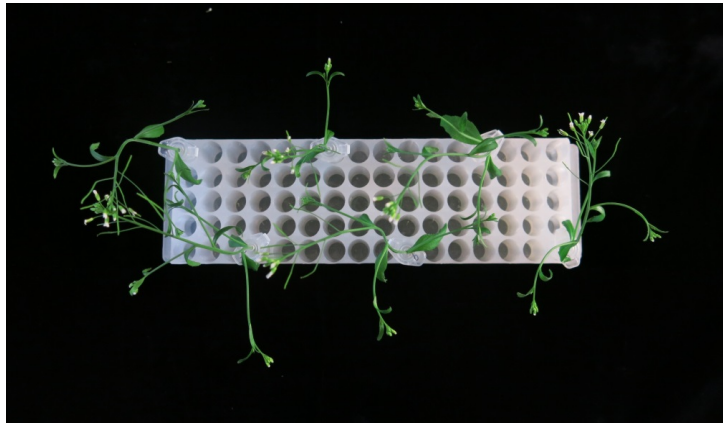


Figure 2.2 Excised stems incubated in tubes with MS medium.

(A) An Arabidopsis stem was excised and placed into a 1.5 mL tube containing liquid MS medium.

(B) Arabidopsis stems incubated in MS medium were placed in a rack to perform feeding experiment (picture taken from side).

(C) Stems were sitting away from each other to mimic their growth in the soil (picture taken from top).

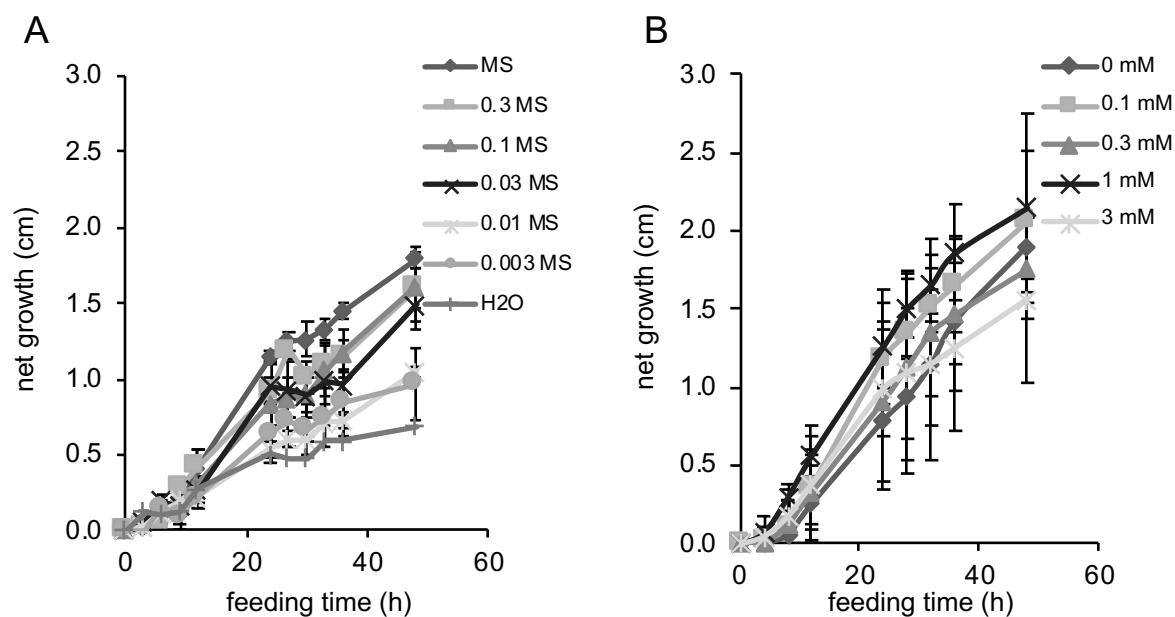


Figure 2.3 Growth of excised Arabidopsis stems in liquid medium.

- (A) Net growth in height of four-week-old Arabidopsis stems excised and incubated in liquid MS medium of different strength for 48 hours. Data represent mean \pm standard deviation (n=3).
- (B) Net growth in height of four-week-old Arabidopsis stems excised and incubated in liquid MS medium containing Phe of different concentrations from 0 mM to 3 mM. Data represent mean \pm standard deviation (n=3).

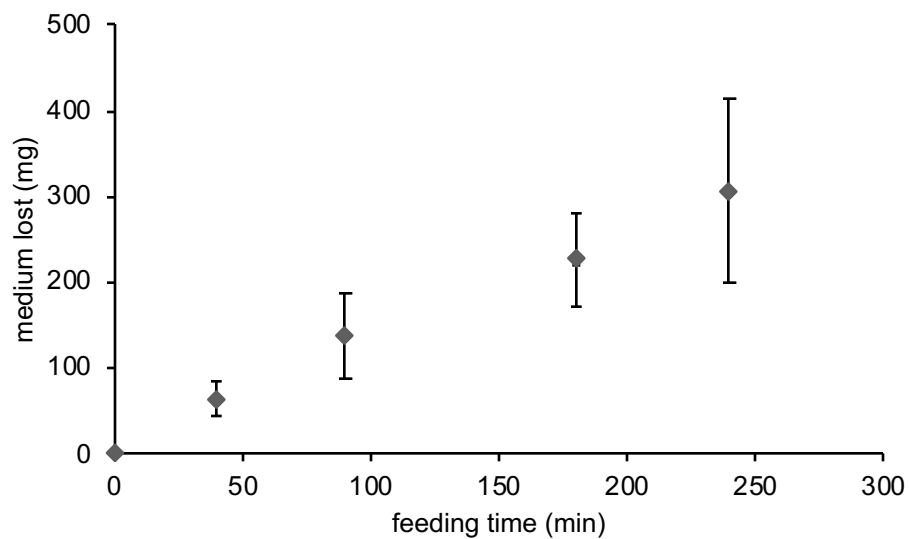


Figure 2.4 Medium absorbed by the excised stems during the feeding process.

The loss of medium from each tube with an excised stem was measured after feeding for 0 min, 40 min, 90 min, 180 min, and 240 min. Data represented mean \pm standard deviation ($n = 45$).

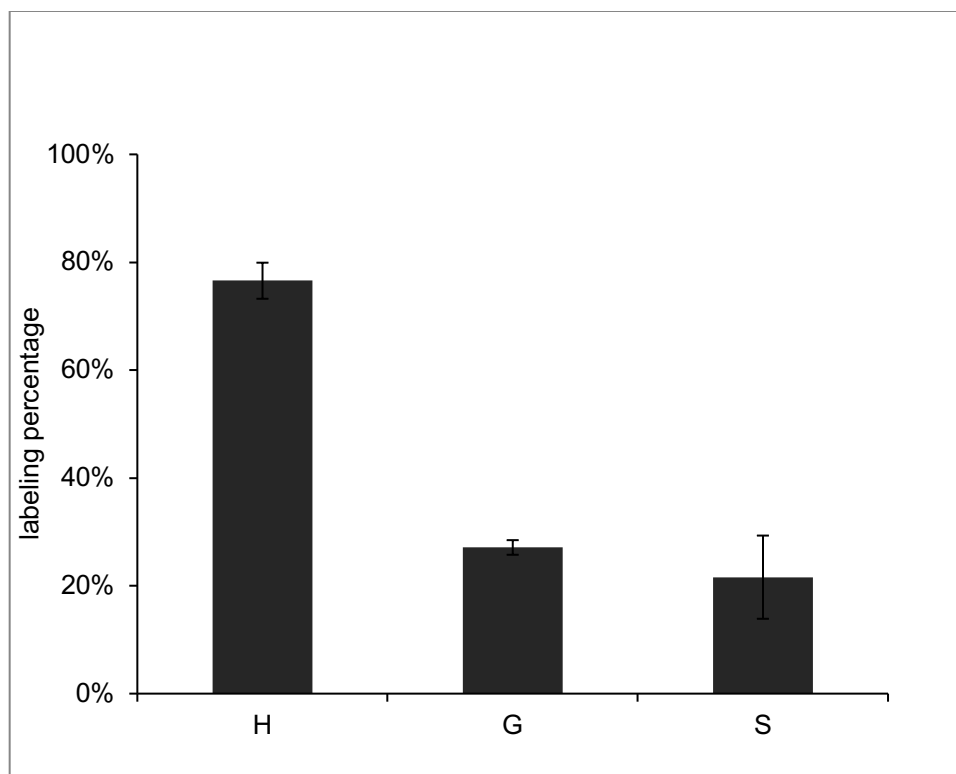


Figure 2.5 Labeling incorporation into lignin in stems fed with $[^{13}\text{C}_6]$ -Phe.

Labeling percentage of lignin monomers in stems fed with 0.3 mM $[^{13}\text{C}_6]$ -Phe for 48 hours were analyzed by DFRC/GC/MS. Data represent mean \pm standard deviation (n=3). The means were compared by one-way ANOVA and statistical difference by Tukey HSD test ($p < 0.01$) was indicated by letters.

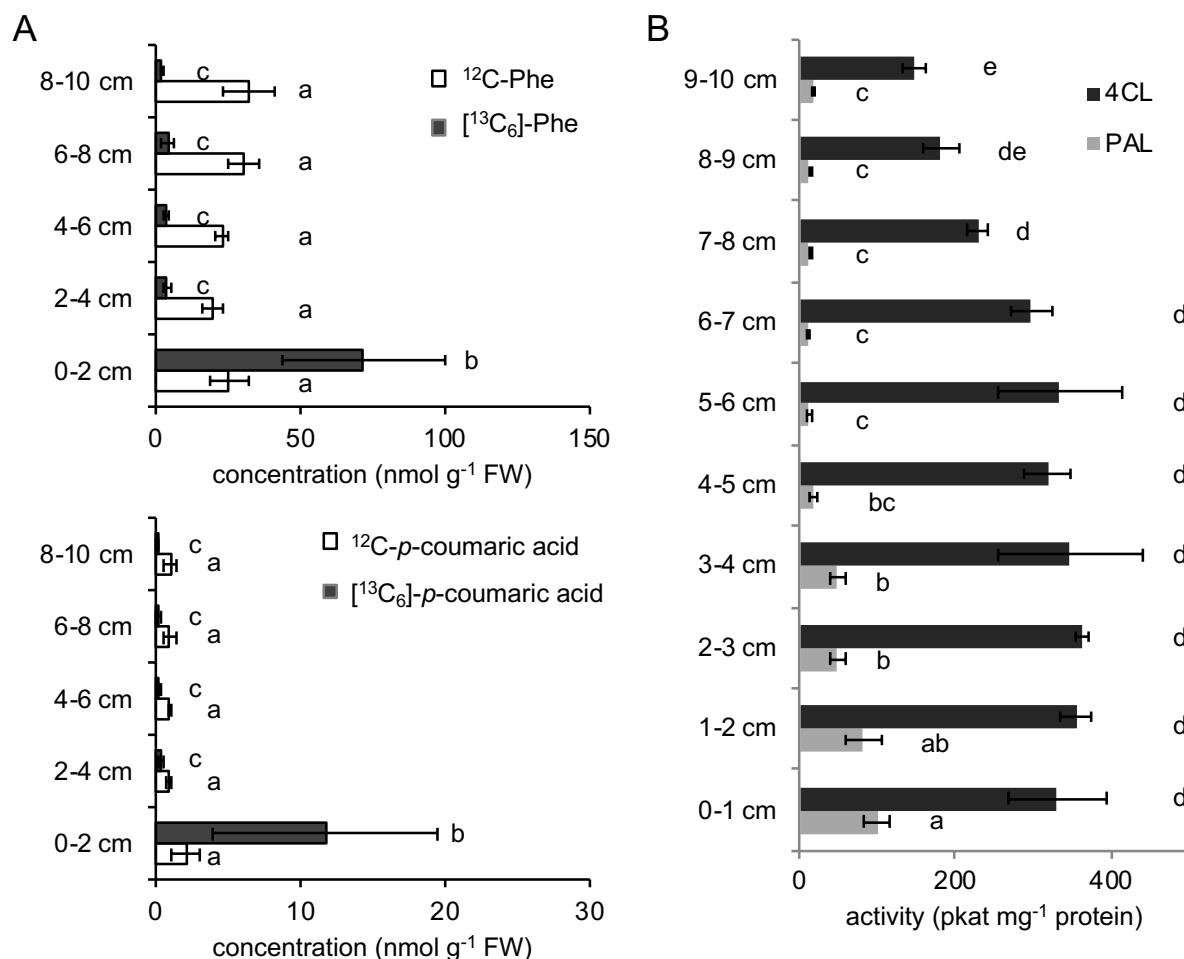


Figure 2.6 Phenylpropanoid metabolism along wild-type Arabidopsis stems.

(A) Content of endogenous (white bars) and labeled (black bars) Phe and p-coumarate along the stems of four-week-old wild-type Arabidopsis fed with 0.3 mM [¹³C₆]-Phe for 6 hours. Data represent mean \pm standard deviation (n=3). One-way ANOVA was used for unlabeled or labeled metabolites and statistical difference by Tukey HSD test ($p < 0.01$) was indicated by letters.

(B) Activities of PAL and 4CL from base to top along the stems of four-week-old Arabidopsis without feeding. Data represent mean \pm standard deviation (n=3). One-way ANOVA was tested for PAL and 4CL assays respectively and statistical difference by Tukey HSD test ($p < 0.01$) was indicated by different letters.

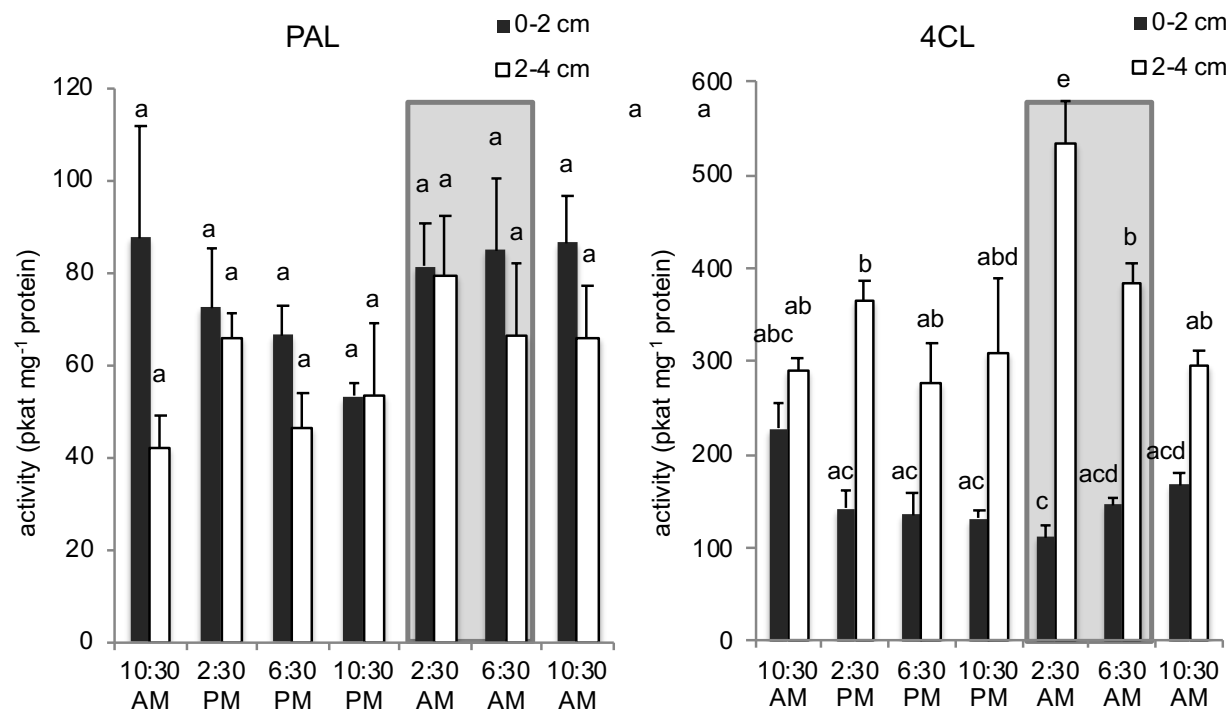


Figure 2.7 PAL and 4CL activities in Arabidopsis stems over a day-night cycle.

Activities of PAL and 4CL from the basal 0-2 cm (black bars) or 2-4 cm (white bars) of four-week-old Arabidopsis stems. Shaded area means stems were sampled at night. Data represent mean \pm SD (n=3). One-way ANOVA was tested for PAL and 4CL assays respectively and statistical difference by Tukey HSD test ($p < 0.01$) was indicated by different letters. Bars with a shared letter have no statistical difference from each other.

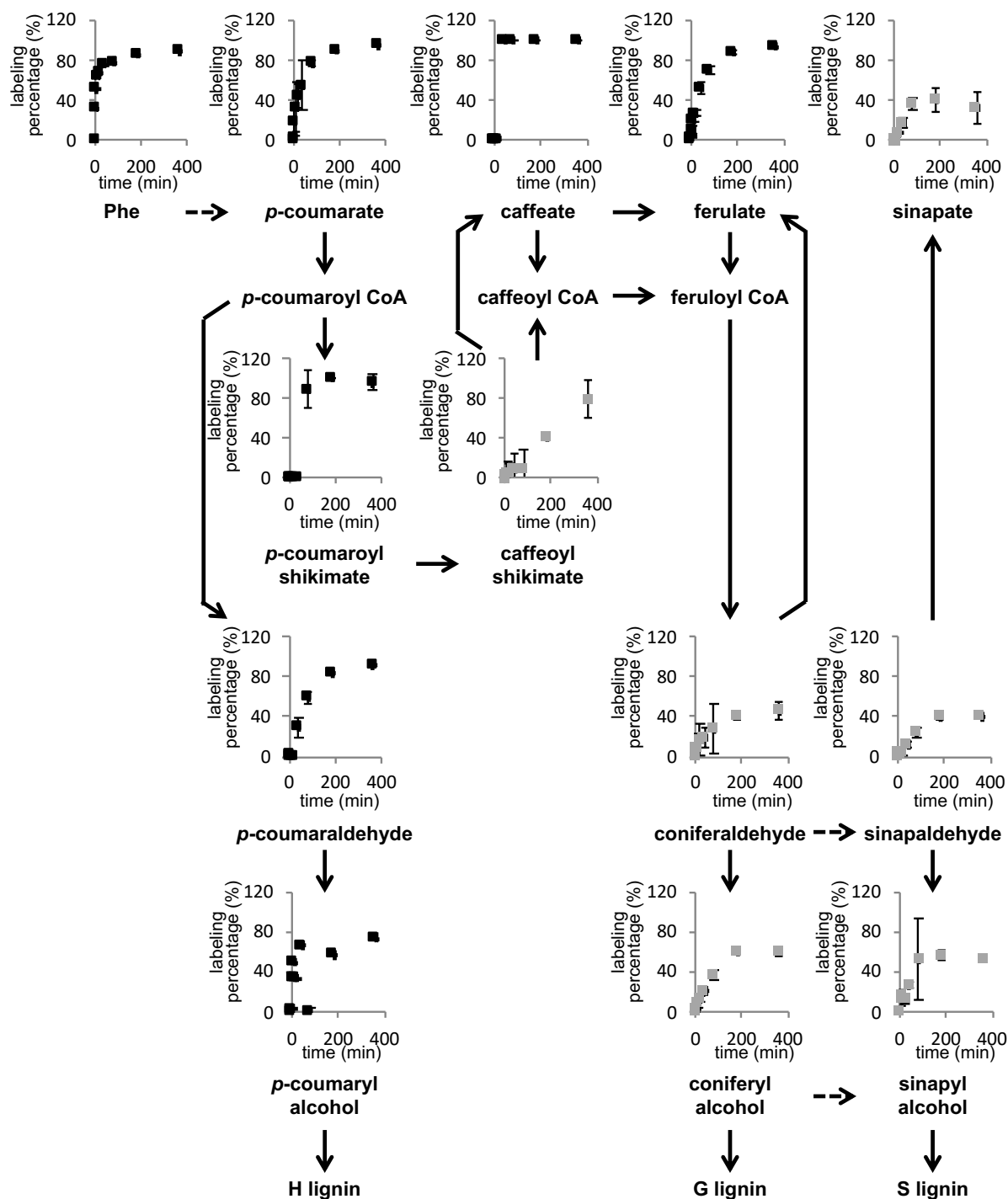


Figure 2.8 Labeling percentage of soluble phenylpropanoids over the feeding time course.

Metabolites were measured from the base of Arabidopsis stems supplied with 1 mM [$^{13}\text{C}_6$]-Phe. The plot of each metabolite was placed above its name on the pathway. Dashed lines mean multiple steps. Data represent mean \pm SD (n=3). Black squares indicate these metabolites were in the first group of the hierarchical clustering in Figure 2.9, and grey squares indicate these metabolites were in the second group.

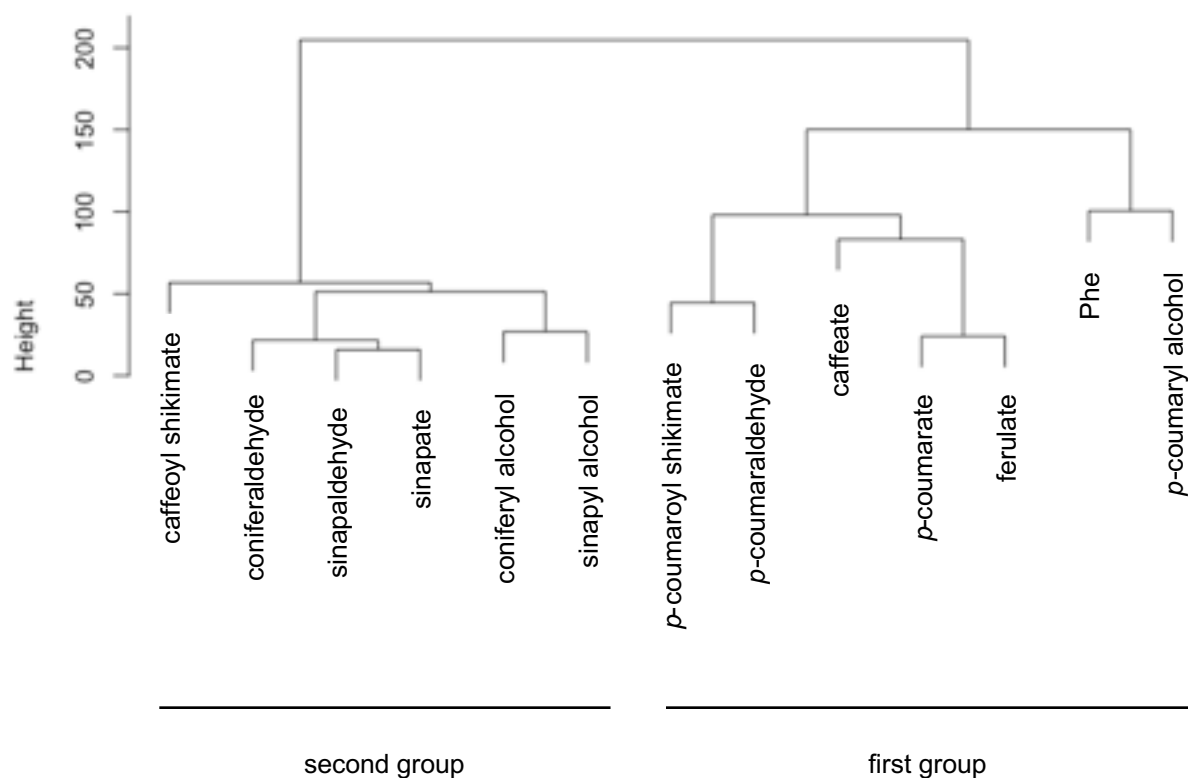


Figure 2.9 Hierarchical clustering of labeling percentage profiles of soluble phenylpropanoids. The labeling enrichment of phenylpropanoids were clustered from the base of Arabidopsis stems supplied with 1 mM [$^{13}\text{C}_6$]-Phe over the feeding time course. The averaged labeling percentage data of each metabolite over the time course from Figure 2.8 were clustered based on squared Euclidian distance.

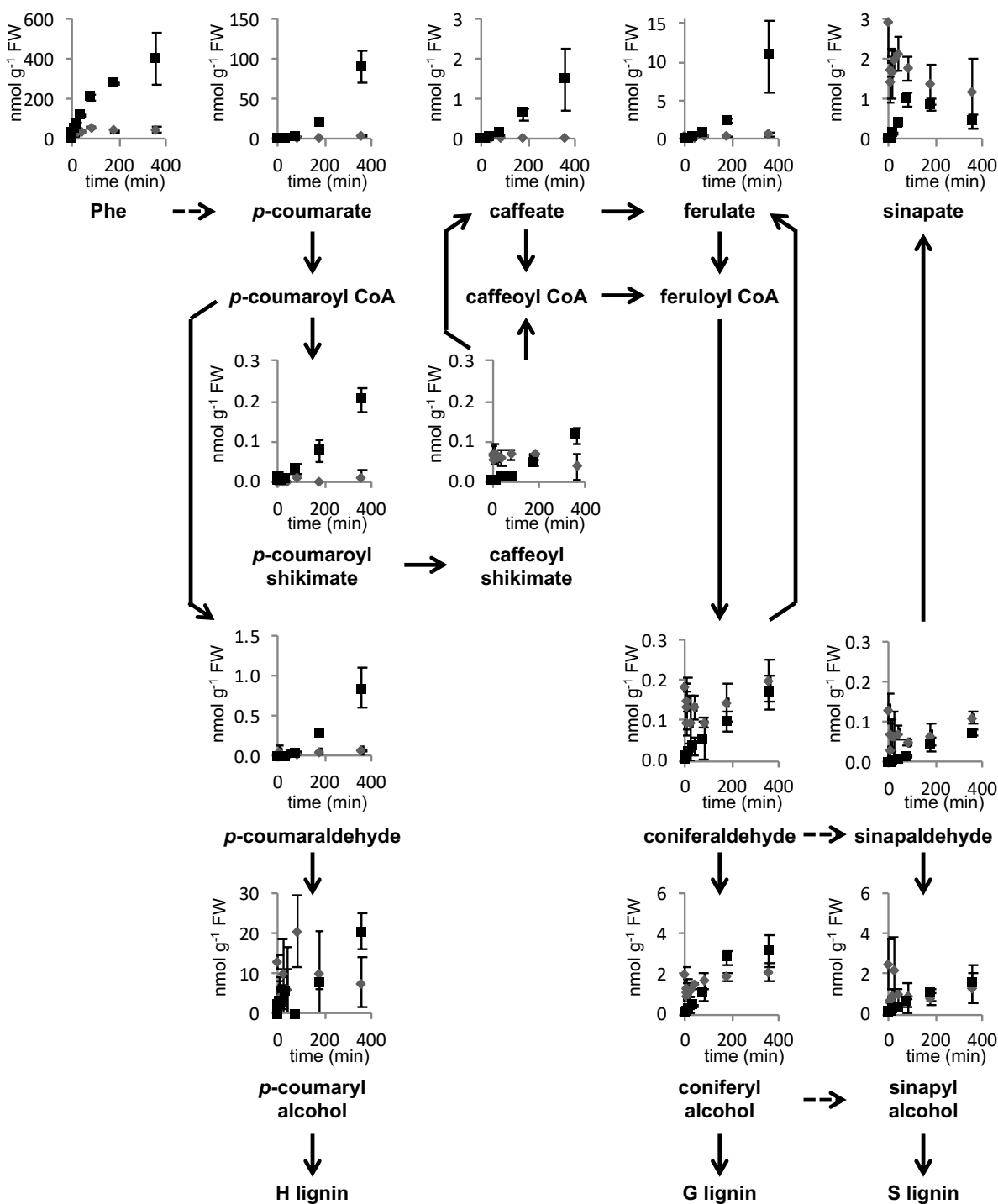


Figure 2.10 Profiles of labeled and unlabeled phenylpropanoids over the feeding time course.

Endogenous (grey) and ¹³C₆ labeled (black) compounds were quantified from the base of *Arabidopsis* stems supplied with 1 mM [¹³C₆]-Phe with LC-MS/MS and normalized to fresh weight of stem tissue. The sum of endogenous and labeled concentrations of each metabolite can be found in Figure 2.11. The plot of each metabolite measured was placed above its name on the pathway. Dashed lines mean multiple steps. Data represent mean ± SD (n=3).

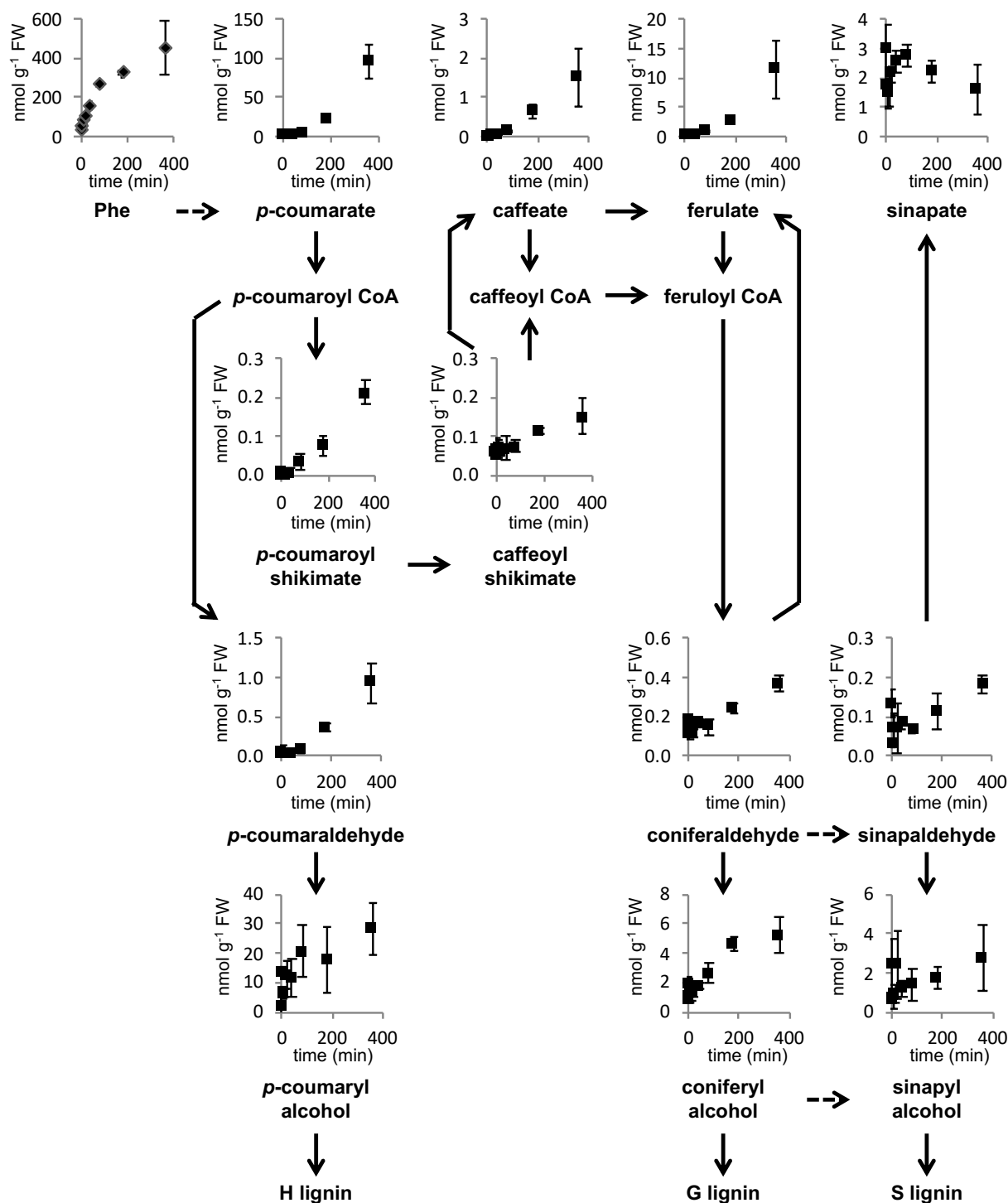


Figure 2.11 Profiles of total phenylpropanoids over the feeding time course.

Sum of endogenous and ¹³C₆ labeled compounds was quantified from the base of Arabidopsis stems supplied with 1 mM [¹³C₆]-Phe with LC-MS/MS and normalized to fresh weight of stem tissue. The plot of each metabolite measured was placed above its name on the pathway. Dashed lines mean multiple steps. Data represent mean ± SD (n=3).

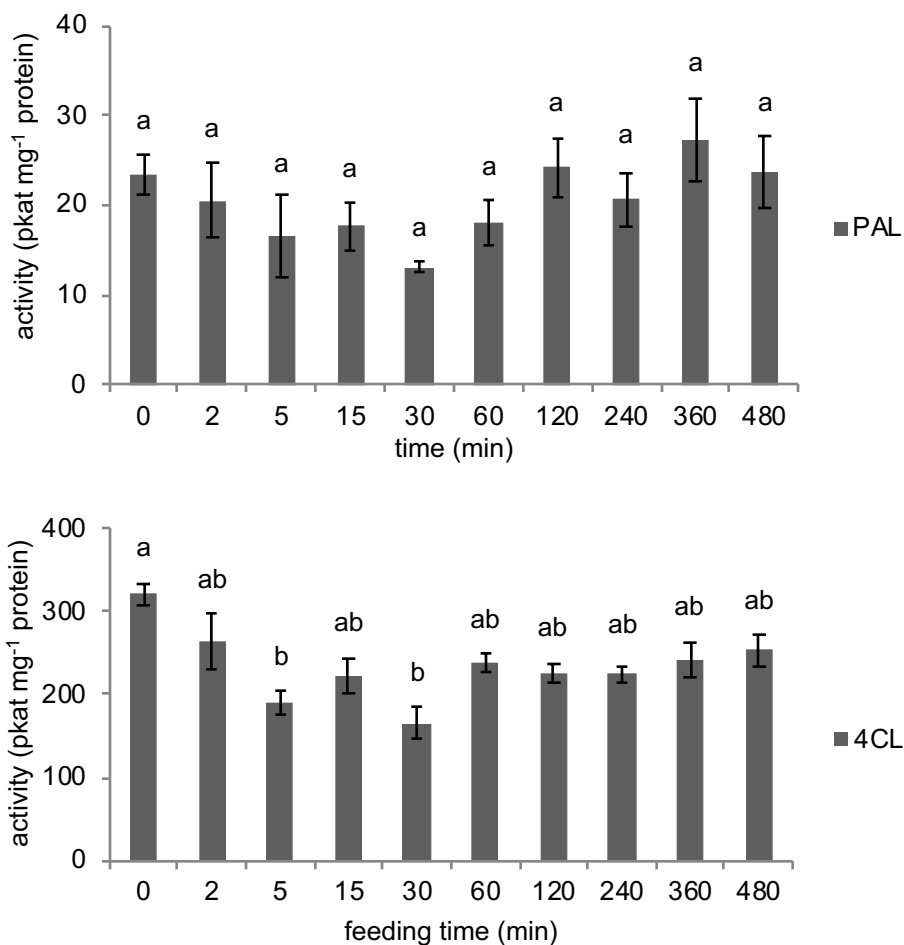


Figure 2.12 PAL and 4CL activities in the stems over the feeding course.

Activities of PAL and 4CL were measured from the basal 0-2 cm of four-week-old *Arabidopsis* stems fed with 1 mM Phe. Data represent mean \pm SD (n=3). One-way ANOVA was tested for PAL and 4CL assays respectively and statistical difference by Tukey HSD test ($p < 0.01$) was indicated by different letters.

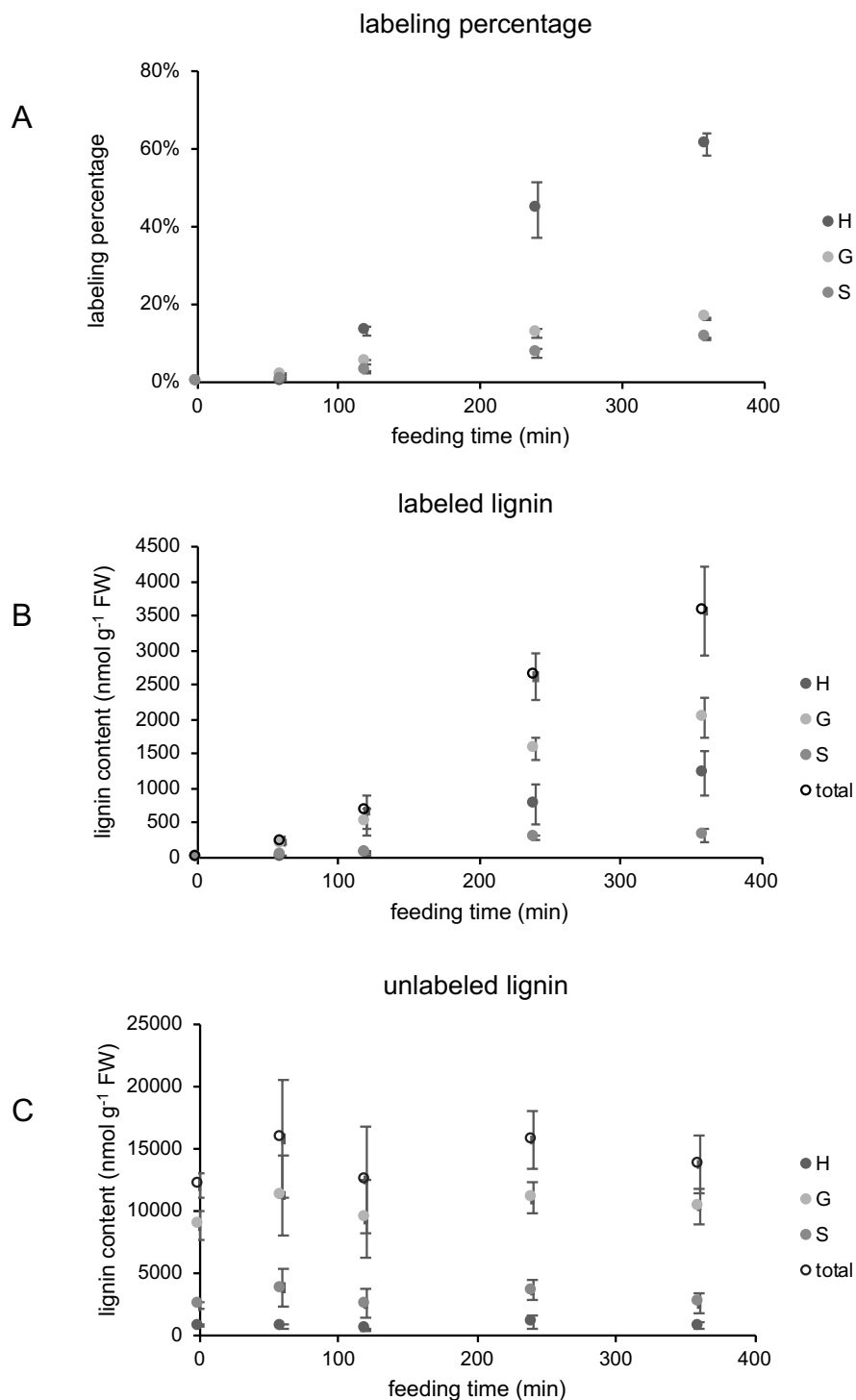


Figure 2.13 Analysis of lignin monomers in excised Arabidopsis stems fed with [$^{13}\text{C}_6$]-Phe.

Lignin was analyzed by DFRC/GC/FID/MS from the base of four-week-old wild-type Arabidopsis stems fed with 1 mM [$^{13}\text{C}_6$]-Phe for 0, 60, 120, 240, or 360 min. Labeling percentage (A) of each monomer was calculated using the labeled monomer (B) divided by the sum of labeled and unlabeled (C). Data represent mean \pm standard deviation ($n=3$).

CHAPTER 3. DYNAMIC MODELING OF SUBCELLULAR PHENYLPROPANOID METABOLISM IN ARABIDOPSIS LIGNIFYING CELLS

3.1 Abstract

Lignin is a polymer that significantly inhibits saccharification of plant feedstocks. Adjusting the composition or reducing the total lignin content have both been demonstrated to result in an increase in sugar yield from biomass. However, because lignin is essential for plant growth, it cannot be manipulated with impunity, thus it is important to understand the control of carbon flux towards lignin biosynthesis such that optimal modifications to it can be made precisely. Phenylalanine (Phe) is the common precursor for all lignin subunits and it is commonly accepted that all biosynthetic steps, spanning multiple subcellular compartments, are known, yet an *in vivo* model of how flux towards lignin is controlled is lacking. To address this deficiency, we formulated and parameterized a kinetic model based on data from feeding *Arabidopsis thaliana* basal lignifying stems with ring labeled [$^{13}\text{C}_6$]-Phe. Several candidate models were compared by an information theoretic approach to select the one that best matched the experimental observations. Here we present a dynamic model of phenylpropanoid metabolism across several subcellular compartments that describes the allocation of carbon towards lignin biosynthesis in wild-type *Arabidopsis* stems. Flux control coefficients for the enzymes in the pathway starting from arogonate dehydratase through 4-coumarate: CoA ligase were calculated and show that plastidial cationic amino-acid transporter has the highest impact on flux.

3.1.1 Keywords

^{13}C isotopic labeling; Kinetic modeling; Model selection; Phenylpropanoid metabolism; Lignin biosynthesis

3.2 Introduction

Plant lignocellulosic biomass is one of the most promising resources for second-generation biofuel production because of its ready availability and reduced environmental impacts [1]. It is estimated that up to 442 billion liters of bioethanol could be produced each year from global lignocellulosic biomass [2], yet commercial application is largely inhibited by the complex

production process that currently render it economically unfavorable [3]. One of the technical barriers lies in inefficient saccharification due to the presence of lignin in plant's secondary cell wall. Lignin impedes the process by either physically restricting cellulose/hemi-cellulose accessibility [4], or adsorbing hydrolytic enzymes such as cellulase [5, 6]. An improvement in hydrolysis efficiency and a reduction in cost is thus expected by either reducing the total amount of lignin, or manipulating its composition [7, 8]. A better understanding of how plant allocates carbon into lignin biosynthesis will be beneficial for these engineering efforts.

The general phenylpropanoid pathway connects primary metabolism with lignin biosynthesis in plants. All monolignols found in the lignin of dicotyledonous plants are derived from Phe, which is the key product of shikimate pathway and precursor of phenylpropanoid compounds (Figure 3.1). Multiple compartments are involved in phenylpropanoid metabolism, but genetic studies showed that Phe is predominantly synthesized inside the plastid, and translocated into cytosol with a cationic-amino acid transporter [9-12]. Moreover, isotopic labeling experiments in petunia flowers revealed that 84.8% of Phe is produced inside the plastid during the period of active volatile production [12].

Plants possess several mechanisms to control flux through the phenylpropanoid network by regulating the synthesis of Phe and the activities of downstream phenylpropanoid enzymes as well as by sequestering pathway intermediates. Arogenate dehydratase (ADT) in plastids, which catalyzes the conversion of arogenate to Phe, is known to be feedback inhibited by its product [13, 14]. Another mechanism includes post-translational regulation over PAL, the gateway enzyme for phenylpropanoid metabolism. The Kelch repeat F-box (KFB) protein family interacts with PAL thus mediating its ubiquitination and degradation [15-17]. In addition, cinnamate, the direct product of PAL, competitively inhibits PAL in several plant species [18-20]. Furthermore, plants utilize compartmental sequestration of pathway intermediates to control the flux through network. Indeed, nearly 90% of *p*-coumarate, the product of C4H was detected in the vacuole of soybean [21]. There is also recent evidence showing the importance of the vacuole in sequestering Phe under metabolic imbalance via a vacuolar Phe transporter [22, 23].

Kinetic modeling is a suitable approach to understand plant metabolism because it provides an integrative and quantitative framework [19, 24-27]. Because of its mechanistic nature, once a valid kinetic model has been developed, metabolic control analysis can be applied to determine which step(s) within the network has the highest control of flux [19, 27, 28]. In addition,

manipulations can be explicitly simulated to predict possible outcomes of pathway perturbation [24, 29]. These predictions are helpful as they provide guidance for metabolic engineering efforts to reduce the number of genetic manipulations for experimental investigation [24].

Accurate model development involves selection of the best model(s) from multiple possible model candidates [30]. Akaike's Information Criteria (AIC) stems from the estimation of relative Kullback-Leibler distance of each model towards the ground truth, which is a fundamental information measure for any models. Thus, it provides an information theoretic way for model selection [31, 32], which considers the balance between model's performance and complexity. Using AIC and an *Escherichia coli* kinetic model, Link *et al.* successfully identified previously unknown allosteric interactions that are likely active *in vivo* in *E. coli* [33].

To date there are several mathematical models of lignin biosynthesis that have been developed by different groups [34-36]. Lee *et al.* and Faraji *et al.* have applied generalized mass action kinetics to simulate each reaction within the pathway, and they observed that the model was able to fit the observations only when specific metabolic channels were considered [34, 36]. On the other hand, Wang *et al.* simulated the pathway with Michaelis-Menten kinetics, and all kinetic parameters were measured by *in vitro* assays [35]. Allosteric inhibitions towards enzymes from pathway intermediates were included in the model as well if the corresponding K_i values could be quantified. However, these models did not consider subcellular compartments and were not able to explore which of the regulatory mechanisms mentioned above contribute the most to control of carbon flux towards lignin biosynthesis. In this work, we developed a multicompartmental kinetic model connecting Phe biosynthesis with lignin production. By applying AIC, we identified a physiologically relevant kinetic model for general phenylpropanoid metabolism in Arabidopsis lignifying cells. The developed dynamic model quantitatively captures the critical regulatory properties of phenylpropanoid metabolism controlling carbon flux towards lignin biosynthesis.

3.3 Materials and Methods

3.3.1 Feeding study in wild type Arabidopsis stem

Arabidopsis thaliana (Columbia-0) wild type plants were grown in soil at 23 °C under 16/8-hour light/dark cycle. Feeding studies were conducted on primary stems of around 10 cm [37]. For each treatment, stems were excised and inserted into the medium with [$^{13}\text{C}_6$]-Phe (0.1,

0.3, 1 and 3 mM). [$^{13}\text{C}_6$]-Phe was purchased from Cambridge Isotope Laboratories, Inc. To avoid injury-induced artefacts due to cutting, only the segments 0.5-2 cm from the base of the stems were harvested, at time points of 0, 40, 90, 120, 180 and 240 min after insertion. Three replicates were collected and immediately frozen in liquid nitrogen for each time point, with 10 stem fractions pooled together as one replicate. Stem tissue was ground in liquid nitrogen with a mortar and pestle. An aliquot of each sample from time points at 0, 40, 90, 240 min were taken for enzyme assays. The rest of the sample was used for soluble phenylpropanoids analysis using LC-MS/MS [38] and lignin quantification by acetyl bromide lignin analysis and derivation followed by reductive cleavage (DFRC)/GC/FID/MS [39]. See sections below for detailed methods. Profiles of Phe and *p*-coumarate were directly applied for kinetic modeling. Sum of [$^{13}\text{C}_6$]-lignin monomers and all downstream intermediates of *p*-coumarate was used to represent the labeled product profile for each treatment.

3.3.2 LC-MS/MS analysis

The LC-MS/MS method developed by Jaini *et al.* [38] was applied to quantify metabolites within the pathway. Briefly, tissue samples were extracted in 75% methanol (10 μL per mg FW) and incubated for 2 hours at 65 °C. The samples were centrifuged for 10 min at 16,000 \times g, and the supernatants were then dried in a speed-vac. Samples were re-dissolved in 60 μL 50% methanol and analyzed using a Zorbax Eclipse C8 column (150 mm 4.6 mm, 5 μm , Agilent Technologies, Santa Clara, CA) as described in Jaini *et al.* [38]. A QTrap 5500 triple quadrupole mass spectrometer (AB Sciex, Redwood City, CA) was used to detect soluble metabolites by multiple reaction monitoring in the negative ion mode. Quantification was achieved with calibration curve generated with standards of Phe, cinnamate, *p*-coumarate, caffeate, ferulate, sinapate, shikimate, *p*-coumaraldehyde, coniferaldehyde, sinapaldehyde, *p*-coumaryl alcohol, coniferyl alcohol, sinapyl alcohol, *p*-coumaroyl shikimate, and caffeoyl shikimate. The same calibration was used to for both unlabeled and labeled isotopologues.

3.3.3 Lignin analysis

After soluble metabolites extraction, the cell wall debris was further washed with 10 mL 70% ethanol (v/v) five times and once with acetone. The cell wall residues were then dried at room temperature for lignin analysis. Total lignin was quantified using acetyl bromide lignin analysis

following [40]. Briefly, about 2 mg cell wall residue was weighed and dissolved in 2.5 mL solution of acetyl bromide: glacial acetic acid (25:75) overnight at room temperature. Samples were mixed with 2.5 mL 2 M NaOH, 12 mL glacial acetic acid, and 0.5 mL 7.5 M hydroxylamine hydrochloride in a 50-mL volumetric flask. The final volume was brought to 50 mL with glacial acetic acid. Absorbance at 280 nm was measured for each sample and total lignin was calculated using the extinction coefficient of $23.29 \text{ g}^{-1} \text{ L cm}^{-1}$. The rest of each cell wall residue sample was weighed and used for monomer composition analysis with DFRC/GC/FID/MS. Samples were dissolved overnight at room temperature in 2.5 mL of a solution of acetyl bromide in acetic acid (20:80) with 0.2 mg 4,4'-ethylidenebisphenol as internal standard. The products were dried then dissolved in 2 mL dioxane: acetic acid: water (50:40:10). 50 mg zinc dust was added into each sample and mixed well. After 20 min of reaction, about 200 μL supernatants were loaded onto a LC-Si solid phase extraction column pre-conditioned with 95% ethanol and water and washed with 2.5 mL 25% ethanol. Samples were eluted with 2.5 mL 95% ethanol then dried before derivatization with 0.5 mL anhydrous acetic anhydride: pyridine (60:40) overnight. Acetylated products were then dried and dissolved in 200 μL dichloromethane. A 1 μL sample was analyzed on GC/FID for monomer composition and on GC/MS for labeling percentage of each monomer. To determine *de novo* lignin deposition rate, we harvested 0.5-2 cm stem sections from soil-grown plants each day for five days after bolting. Total lignin content in the sections was quantified with acetyl bromide method. Linear regression was applied to estimate the slope, which was used to determine the *de novo* lignin deposition rate.

For labeled lignin quantification at each time point, we first measured the composition of each monomer released from lignin with DFRC/GC/FID. Then we estimated the amount of each monomer by multiplying total acetyl bromide lignin content. The labeled lignin content was calculated by multiplying the total amount of each monomer by its labeling percentage determined from DFRC/GC/MS analysis. The sum of all three types of labeled lignin was used to estimate the labeled lignin deposition rate.

3.3.4 Enzyme assays

Crude proteins were extracted from each sample with Tris-HCl buffer (pH 7.8) and desalted. PAL assays were performed in 100 mM Tris-HCl buffer (pH 7.5) with 5 mM Phe and 5 μL protein extract in a final volume of 50 μL , by incubating at 23 °C for 120 min. 4CL assays

were performed in 100 mM Tris-HCl buffer (pH 7.5) with 5 mM MgCl₂, 5 mM ATP, 1 mM *p*-coumarate, 0.3 mM CoA and 2 µL protein extract in a final volume of 40 µL. Reactions were incubated at 23 °C for 20 min. Both assay products were analyzed with HPLC on a Shim-pack XR-ODS column (Shimadzu; column dimensions, 3 x 75 mm) using acetonitrile and 0.1% formic acid as mobile phases, and quantified using cinnamate and *p*-coumaroyl CoA as standards. Protein concentrations were determined by Bradford assay with bovine serum albumin as standard.

3.3.5 Kinetic modeling

The pathway is assumed to be at a metabolic steady state prior to external Phe feeding. The output flux from the pathway is set equal to the average lignin deposition rate (3.44 nmol g FW⁻¹ min⁻¹) measured from 5-week-old Arabidopsis basal stems, as lignin is the major sink for the phenylpropanoid pathway in this type of tissue. The refined model also has two extra fluxes towards vacuole storage for both Phe and *p*-coumarate. The model's stoichiometry is shown in Figure 3.2 with abbreviations detailed in the caption. The mass balance equations for the refined model are:

$$\frac{d[^{12}\text{Phe}_{\text{plastid}}]}{dt} = {}^{12}v_{ADT} - {}^{12}v_{PCAT} \quad (1)$$

$$\frac{d[^{12}\text{Phe}_{\text{cytosol}}]}{dt} = {}^{12}v_{PCAT} - {}^{12}v_{PAL} - {}^{12}v_{VCAT} \quad (2)$$

$$\frac{d[^{13}\text{Phe}_{\text{cytosol}}]}{dt} = {}^{13}v_{PXT} - {}^{13}v_{PAL} - {}^{13}v_{VCAT} \quad (3)$$

$$\frac{d[^{12}\text{Phe}_{\text{vacuole}}]}{dt} = {}^{12}v_{VCAT} \quad (4)$$

$$\frac{d[^{13}\text{Phe}_{\text{vacuole}}]}{dt} = {}^{13}v_{VCAT} \quad (5)$$

$$\frac{d[^{12}\text{pCA}_{\text{cytosol}}]}{dt} = {}^{12}v_{PAL} - {}^{12}v_{4CL} - {}^{12}v_{PVT} \quad (6)$$

$$\frac{d[^{13}\text{pCA}_{\text{cytosol}}]}{dt} = {}^{13}v_{PAL} - {}^{13}v_{4CL} - {}^{13}v_{PVT} \quad (7)$$

$$\frac{d[^{12}\text{pCA}_{\text{vacuole}}]}{dt} = {}^{12}v_{PVT} \quad (8)$$

$$\frac{d[^{13}\text{pCA}_{\text{vacuole}}]}{dt} = {}^{13}v_{PVT} \quad (9)$$

$$\frac{d[^{13}\text{Product}]}{dt} = {}^{13}v_{4CL} \quad (10)$$

$$[^{13}\text{Phe}_{\text{xylem}}] = \text{Volume}_{\text{xylem}} * [^{13}\text{Phe}_{\text{medium}}] \quad (11)$$

For the base model, mass balances of vacuolar metabolites and corresponding vacuolar transporter reactions are absent. For the models with cinnamate and/or *p*-coumaroyl CoA introduced, similar mass balance equations are introduced for both ^{12}C and ^{13}C isotopes. The Phe concentration in the xylem is assumed to be equal with the Phe concentration in the feeding medium. Essentially this means the transpiration rate is assumed to be much faster than cellular uptake, so that xylem is instantaneously filled with Phe upon feeding, and the concentration remains constant throughout the feeding period. The rest of reactions are modeled with Michaelis-Menten equations, with each isotopic molecule competing for the same enzymatic sites. Detailed reaction kinetics in the refined model are listed below:

$$^{12}v_{ADT} = \frac{V_{ADT}^{apparent}}{1 + \frac{[^{12}\text{Phe}_{plastid}]}{K_{i,Phe}^{ADT}}} \quad (12)$$

$$^{12}v_{PCAT} = V_{max,PCAT} \frac{[^{12}\text{Phe}_{plastid}]}{K_{M,Phe}^{PCAT} \left(1 + \frac{[^{12}\text{Phe}_{cytosol}] + [^{13}\text{Phe}_{cytosol}]}{K_{i,Phe}^{PCAT}}\right) + [^{12}\text{Phe}_{plastid}]} \quad (13)$$

$$^{13}v_{PXT} = V_{max,PXT} \frac{[^{13}\text{Phe}_{xylem}]}{K_{M,Phe}^{PXT} + [^{13}\text{Phe}_{xylem}]} \quad (14)$$

$$^{12}v_{PAL} = V_{max,PAL} \frac{[^{12}\text{Phe}_{cytosol}]}{K_{M,Phe}^{PAL} + [^{12}\text{Phe}_{cytosol}] + [^{13}\text{Phe}_{cytosol}]} \quad (15)$$

$$^{13}v_{PAL} = V_{max,PAL} \frac{[^{13}\text{Phe}_{cytosol}]}{K_{M,Phe}^{PAL} + [^{12}\text{Phe}_{cytosol}] + [^{13}\text{Phe}_{cytosol}]} \quad (16)$$

$$^{12}v_{VCAT} = V_{max,VCAT} \frac{[^{12}\text{Phe}_{cytosol}]}{K_{M,Phe}^{VCAT} + [^{12}\text{Phe}_{cytosol}] + [^{13}\text{Phe}_{cytosol}]} \quad (17)$$

$$^{13}v_{VCAT} = V_{max,VCAT} \frac{[^{13}\text{Phe}_{cytosol}]}{K_{M,Phe}^{VCAT} + [^{12}\text{Phe}_{cytosol}] + [^{13}\text{Phe}_{cytosol}]} \quad (18)$$

$$^{12}v_{4CL} = V_{max,4CL} \frac{[^{12}p\text{CA}_{cytosol}]}{K_{M,pCA}^{4CL} + [^{12}p\text{CA}_{cytosol}] + [^{13}p\text{CA}_{cytosol}]} \quad (19)$$

$$^{13}v_{4CL} = V_{max,4CL} \frac{[^{13}p\text{CA}_{cytosol}]}{K_{M,pCA}^{4CL} + [^{12}p\text{CA}_{cytosol}] + [^{13}p\text{CA}_{cytosol}]} \quad (20)$$

$$^{12}v_{PVT} = V_{max,PVT} \frac{[^{12}p\text{CA}_{cytosol}]}{K_{M,pCA}^{PVT} + [^{12}p\text{CA}_{cytosol}] + [^{13}p\text{CA}_{cytosol}]} \quad (21)$$

$$^{13}v_{PVT} = V_{max,PVT} \frac{[^{13}p\text{CA}_{cytosol}]}{K_{M,pCA}^{PVT} + [^{12}p\text{CA}_{cytosol}] + [^{13}p\text{CA}_{cytosol}]} \quad (22)$$

When the mass balances on either cinnamate or *p*-coumaroyl CoA are considered, the reaction kinetics for C4H or CCR is introduced as a one-substrate Michaelis-Menten equation.

Also with the presence of *p*-coumaroyl CoA, free CoA is modeled as an independent substrate for 4CL kinetics, which leads to the new equation for 4CL reaction:

$$^i v_{4CL} = V_{max,4CL} \frac{[^i pCA_{cytosol}][CoA]}{K_{M,pCA}^{4CL} K_{M,CoA}^{4CL} + (K_{M,CoA}^{4CL} + [CoA]) * ([^{12} pCA_{cytosol}] + [^{13} pCA_{cytosol}])}, \quad i = 12 \text{ or } 13 \quad (23)$$

In this case, net synthesis of CoA during the feeding study is assumed to be negligible, so that its mass balance is:

$$\frac{d[CoA]}{dt} = ^{12} v_{CCR} + ^{13} v_{CCR} - ^{12} v_{4CL} - ^{13} v_{4CL} \quad (24)$$

Each model with potential metabolite-enzyme interactions includes one possible regulatory interaction (activation, competitive or uncompetitive inhibition). The general form of such interactions in the corresponding enzymatic kinetics was modeled as:

$$v = V_{max} \frac{[activator]}{K_a + [activator]} \frac{[substrate]}{K_M * (1 + \frac{[inhibitor]}{K_{ic}}) + [substrate] * (1 + \frac{[inhibitor]}{K_{iu}})} \quad (25)$$

3.3.6 Parameter identification

Not all model parameters are independent as steady state assumption is made at time 0 for the model. For the refined model, the number of independent parameters estimated is 15. Parameters were estimated by minimizing a pre-defined objective function score defined as the sum of square differences between model's predictions and experimental observations weighted by measurement errors given training datasets:

$$J = \sum_{i=1}^q \sum_{t=1}^T \sum_{m=1}^M \frac{(Y_{m,t,i}^{predicted} - Y_{m,t,i}^{measured})^2}{s_{m,t,i}^2} \quad (26)$$

With M available measurements each having T time points, and q different treatment datasets are used for model training.

The objective function was minimized first with *lsqnonlin* in MATLAB (R2013a, Mathworks, Inc), a gradient-based optimization algorithm with a multi-start approach using random initial guesses. The best result was then used as the starting point for the following Markov Chain Monte Carlo (MCMC) sampling algorithm [41] to further explore the global minimum in the parameter space. For each MCMC batch, 100,000 samples were generated for an efficient exploration. The sampling was continued until the objective function score was not reduced with an additional MCMC batch.

For each parameter, the 95% confidence intervals were estimated by ranking 100,000 samples from highest to lowest and extracting values at 2,500th and 97,500th.

3.3.7 Model comparison with information criteria

Models in the current study have different numbers of parameters, so that a direct comparison of the objective function fitting score is not appropriate since it is biased toward models with more parameters. Thus, AIC was applied to compensate for this issue [32]. The form of AIC score corrected for small sample/parameter ratio [32] is:

$$AIC = n \ln \frac{\widehat{sse}}{n} + 2k + \frac{2k(k+1)}{n-k-1} \quad (27)$$

where \widehat{sse} is the sum of squared errors, which is the objective function score; n represents number of independent measurements, and k represents number of free parameters. A lower AIC score indicates an overall better model performance corrected with its complexity.

3.3.8 Metabolic control analysis

Metabolic control analysis provides a quantitative evaluation of the perturbation of enzyme amount towards pathway flux for each metabolic step [29]. The flux control coefficients (FCC) of an enzyme is mathematically defined as:

$$FCC_i = \frac{\frac{\partial J}{J}}{\frac{\partial [e_i]}{[e_i]}} \quad (28)$$

Here J represents flux through the pathway, and $[e_i]$ represents the target enzyme amount. Computationally, starting with a steady state model, V_{max} of the target enzyme was changed $\pm 5\%$ to get to new steady states where new flux values were recorded which were used to compute FCCs.

3.4 Results

3.4.1 Metabolic profiling of wild-type Arabidopsis basal stem fed with ring labeled [$^{13}\text{C}_6$]-Phe

A feeding system in Arabidopsis stem was previously developed [37] using the basal portions of Arabidopsis stems in which the tissue segments were able to take up [$^{13}\text{C}_6$]-Phe from external medium, and incorporate it into downstream phenylpropanoids including lignin. Four feeding experiments were conducted using different [$^{13}\text{C}_6$]-Phe concentrations (0.1, 0.3, 1 and 3 mM) and metabolite profiling was performed at several points over a four-hour period (Figure 3.2). The profiles describe the overall abundances of metabolites within the selected tissue, thus lack

the subcellular details. On the other hand, significant labeling of extracted Phe was already achieved after 40 min of feeding with the degree of labeling increasing proportionally to the concentration of fed [$^{13}\text{C}_6$]-Phe (Figure 3.2). After 40 min, labeled *p*-coumarate and lignin were readily detected in extracts from the stem tissue and the cell wall residue, respectively (Figure 3.2 and Figure 3.3), suggesting that the exogenous [$^{13}\text{C}_6$]-Phe is transported into stem tissue and efficiently incorporated into pathway end products.

To test if the activities of biosynthetic enzymes remain unchanged during the feeding process, the activities of PAL and 4CL in the stem tissue were determined after 0 min, 40 min, 90 min, and 240 min of feeding. As shown in Supplementary Figure 3.2, there were no changes in PAL and 4CL activities during the feeding process, independent of the exogenous Phe concentrations.

3.4.2 Base kinetic model construction and parameterization

A base kinetic model for the general phenylpropanoid pathway was constructed on the known structure of phenylpropanoid metabolism in Arabidopsis [42]. A putative xylem transporter (PXT) was added to model Phe uptake from the xylem (Figure 3.3). At the same time, the plastidial cationic amino-acid transporter (PCAT) was incorporated in the model to account for unlabeled Phe that is synthesized in the plastid by ADT, and subsequently exported by transporter into the cytosol [12] where Phe is converted to cinnamate by PAL. As cinnamate was below the detection limit for all feeding experiments, it was not included in the base model. Therefore, initially PAL and C4H enzymatic steps were lumped together into one single reaction. The product of this PAL/C4H reaction is *p*-coumarate, which is then consumed by 4CL *en route* to all monolignols in lignin biosynthesis. The base model treats this step as a one-substrate reaction due to lack of measurements of the other reactants, CoA and ATP. PAL and 4CL isoforms were not considered in the current model, as only total activity for these enzymes were measured. Therefore, all V_{\max} and K_m values within the model are apparent ones. In addition to the enzymatic steps directly involved in phenylpropanoid metabolism, two feedback inhibition mechanisms were also included: (i) plastidial Phe competitive inhibition of ADT and (ii) cytosolic Phe inhibition of PCAT (Figure 3.4). There were 11 parameters to estimate in the base model.

An efficient MCMC sampling algorithm was applied for parameter estimation [41]. To evaluate how well the proposed base kinetic model can explain the *in vivo* metabolite dynamics,

three datasets with different levels of fed [$^{13}\text{C}_6$]-Phe (0.1, 0.3 and 3 mM) were first fit with the objective function to train the model, with the 1 mM dataset left out for model validation. Metabolite profiles predicted with estimated parameter values by MCMC for the base model were compared against experimental data (Figure 3.2, dashed lines). Although dynamics of Phe and downstream products were accurately captured by the model, it failed to predict *p*-coumarate trends in all cases (for both training and validation datasets), indicating that the base model was unable to fully capture the experimental observations. As the structure of the network included in the base model is experimentally supported in Arabidopsis [43-47], it is more likely that some regulatory mechanisms active *in vivo* remain unknown and thus had not been considered.

3.4.3 Incorporation of vacuolar storage of Phe and *p*-coumarate significantly improve model performance

Lynch *et al.* identified a vacuole storage mechanism for Phe via a VCAT in petunia [23]. Arabidopsis contains four homologues encoding vacuolar membrane amino acids transporters [48, 49]. Additionally, subcellular fractionation studies have detected the vacuolar pools of Phe [22] and *p*-coumarate [21] *in planta*. Therefore, we hypothesized that the vacuole storage mechanisms are also present in Arabidopsis lignifying cells. The putative vacuolar transporters could remove excessive Phe and *p*-coumarate from the cytosol into this separate compartment, to maintain lignin biosynthetic homeostasis if the concentration of either compound changes substantially. To test this hypothesis, we added two vacuolar transporters into the base model. Because simulation had been performed over a short time period (4 hour), the vacuole transporters were modeled to function unidirectionally from cytosol to vacuole. Thus, each transporter can be modeled with two additional parameters (See Methods for detailed equations). Three competing models were formulated with vacuole storage mechanisms for Phe and/or *p*-coumarate (Figure 3.5). To evaluate which, if any, of these competing models has better performance, the AIC corrected for low sample/parameter ratio was applied to all four datasets. The model with storage mechanisms for both, Phe and *p*-coumarate, had the largest reduction in AIC score relative to the base model, indicating that having those additional transporters significantly improved the fit (Figure 3.5).

To evaluate whether the refined model has prediction capability outside of training data, the refined model with two storage mechanisms was trained using three datasets (0.1, 0.3 and 3 mM), while the 1 mM dataset was used for model validation, just as it was done for the base model. As expected, the refined model better matched the data trends in both the training and validation

sets compared to the base model (Figure 3.2), especially on the predictions of *p*-coumarate profiles. The simulations of the refined model suggest that *p*-coumarate mainly accumulates inside the vacuole during feeding studies (Table 3.1), providing the evidence of an active sequestration mechanism when Phe was present at high concentrations.

Next, all four datasets were utilized for the final parameter identification. Estimated parameters were compared with literature values when available (Table 3.2). The model accurately predicted PAL K_m for Phe which within a 95% confidence interval was consistent with *in vitro* measurements (3.3 versus 4.3 nmol g FW⁻¹). However, the 4CL K_m for *p*-coumarate was largely underestimated (0.30 versus measured 2.4 nmol g FW⁻¹). Since 4CL step was modeled as a single-substrate reaction, it should be noted that the estimated K_m is a combined one including the influence of CoA and ATP. A significantly lower estimated value might indicate the either CoA and/or ATP pool *in vivo* is not saturating 4CL, as occurs in *in vitro* assays. Consistently, 4CL maximal capacity predicted by the model was significantly lower than 4CL activity measurements from crude extracts (Table 3.2 and Figure 3.6), indicating that 4CL was not functioning at its *in vitro* maximal velocity likely due to low CoA and/or ATP concentration.

3.4.4 No *in silico* evidence for existence of other metabolite-enzyme interactions

The proposed models (both base and refined ones) included two competitive inhibitions based on the previous knowledge of the pathway [12-14]. These models have the risk of being biased against unknown factors because *a priori* no other metabolite-enzyme regulatory interactions were considered. The wide range of perturbations undertaken in the feeding studies (from 0.1 up to 3 mM treatment) provides an excellent dataset to search for any significant regulatory interactions. Starting with the refined model, 14 possible metabolite-enzyme interactions were evaluated (Figure 3.4, Table 3.3). These included all possible combinations between metabolite and enzymes interactions: competitive inhibition, uncompetitive inhibition and feedforward activation of enzymes by metabolites. The AIC score was applied to evaluate if a model with any of the proposed interactions had a better performance than the refined model. Since no reduction in AIC score was observed for any model, the data do not support the *in vivo* existence of any tested metabolite-enzyme interactions (Table 3.3).

Expanding the model with additional metabolites did not improve model's performance

Cinnamate was not initially included as an independent metabolite in the model because its amount was below detection limits with LC-MS/MS [38]. Yet it is possible that its presence could have significant impact on the model's behavior, considering that cinnamate inhibits PAL activity in plants including *Arabidopsis* [18]. Additionally, CoA, a substrate for 4CL-catalyze reaction was also not included in the model because of lack of information about its subcellular levels. The product of 4CL is *p*-coumaroyl CoA, which is a precursor for either CCR *en route* to the synthesis of H lignin, or HCT towards production of G and S lignin (Figure 3.1). Both enzymes release the CoA moiety back to the free CoA pool. Since underestimation of the 4CL K_m for *p*-coumarate suggested the possible limitation of CoA *in vivo*, dynamics of CoA pools might play a role in regulating flux through 4CL step. To test the effect of these metabolites, three additional models including the mass balances of cinnamate and/or *p*-coumaroyl CoA (the latter also includes CoA balance) were constructed and fit to the data from all four feeding experiments for comparison against the refined model. As shown in Figure 3.7, including a cinnamate mass balance into the model did not improve the statistical fit. This suggests that C4H is not a limiting step even during feeding experiments with high concentrations of Phe. Incorporation of the *p*-coumaroyl CoA mass balance also did not improve model's performance, suggesting that within those feeding treatments CoA balance did not have significant impact on 4CL flux (Figure 3.7).

3.4.5 Refined model unravels subcellular pathway dynamics during feeding studies

All attempts described above to improve the model revealed that the kinetic model with vacuole storage mechanism for Phe and *p*-coumarate (refined model) best represents *in vivo* situations. This model was then used to quantify two Phe input fluxes via PXT and ADT and the output flux through 4CL under different feeding conditions (Figure 3.8). The [$^{13}\text{C}_6$]-Phe uptake rate from the xylem (via PXT) into the lignifying cells in the basal stems continuously increased with higher fed Phe concentration up to $12 \text{ nmol g FW}^{-1} \text{ min}^{-1}$, thus contributing from 24% up to 89% of flux that reaches 4CL at 0.1 mM and 3 mM of fed Phe, respectively (Figure 3.8). On the other hand, accumulation of cytosolic Phe at high fed Phe concentrations resulted in a decrease in the plastidial Phe synthesis rate (Figure 3.8 and Table 3.1).

Analysis of model-predicted subcellular distribution of Phe and *p*-coumarate revealed that at low Phe fed concentrations (0.1 mM and 0.3 mM) the enzymes in the cytosol are capable of efficiently utilizing increasing amounts of substrates, as both Phe and *p*-coumarate were around or

below their respective K_m for PAL and 4CL, respectively (Table 3.1 and Table 3.2). Vacuole storage mechanisms played important roles for pathway homeostasis since they sequester a significant amount of Phe and *p*-coumarate (29% and 90% out of total for each, at 0.3 mM fed Phe) (Table 3.1). However, at higher fed Phe concentrations (1 mM and 3 mM) the model predicted that the vacuolar transporters became saturated, resulting in dramatic expansion in the cytosolic Phe and *p*-coumarate pools. After 240 min, cytosolic Phe and *p*-coumarate pools were 16.4-fold and 1.8-fold higher respectively at 3 mM fed Phe than at 1 mM fed Phe, suggesting that PAL and 4CL were no longer able to efficiently convert these substrates into downstream products. The hyper-accumulated cytosolic Phe can feedback inhibit PCAT leading to the build-up of plastidial Phe, which in turn inhibits ADT. Interestingly, while an accumulation of Phe was observed in the cytosol (17.5-fold), the plastid Phe level for the 3 mM fed Phe only increased by 1.6-fold than 1 mM fed Phe, mainly due to tight feedback regulation of ADT (Table 3.1).

3.4.6 Metabolic control analysis determines the relative enzymatic control of flux in the general phenylpropanoid pathway

The FCC determined by metabolic control analysis (MCA) provide a quantitative view of the distribution of control that the individual enzymes impose on each flux within the metabolic network. Under non-fed Phe conditions, the highest control coefficients were found upstream of PAL (Table 3.4). Specifically, PCAT had the highest control coefficient over the flux, which is consistent with the role of petunia plastidial cationic amino-acid transporter (*PhpCAT*) in phenylpropanoid metabolism in flowers [12]. Indeed, a 75-80% downregulation of *PhpCAT* transcript levels led to 20-42% reduction in the phenylpropanoid pathway flux. MCA also indicates that downstream enzymes do not have significant controls over the pathway flux in *Arabidopsis* stems under non-fed conditions. When PAL and 4CL activities were measured in crude extracts of basal stem tissue used in the feeding studies, both values were significantly higher than lignin deposition rate (on average over all feeding conditions 10.5 nmol g FW⁻¹ min⁻¹ for PAL, 82.4 nmol g FW⁻¹ min⁻¹ for 4CL, and 3.4 nmol g FW⁻¹ min⁻¹ for lignin synthesis; Table 3.2 and Figure 3.6), suggesting that neither enzyme was functioning at their V_{max} levels. Together with the very low estimated FCCs, these observations suggest that phenylpropanoid metabolism is mainly controlled by Phe biosynthetic reactions and its export from the plastid.

3.5 Discussion

This study developed multiple kinetic models of the phenylpropanoid metabolism in *Arabidopsis* lignifying cells by evaluating different regulatory mechanisms. As plant metabolism is highly compartmentalized [50], subcellular compartments were also integrated in the model allowing estimation of subcellular concentrations of metabolites. Cascade feedback inhibition of plastidial ADT by cytosolic Phe via PCAT was incorporated in the current kinetic model of *Arabidopsis* phenylpropanoid metabolism. The refined model with this inhibition was able to accurately capture Phe profiles in all feeding datasets. Recently it has been suggested that the plant is able to sense the amount of cytosolic Phe and control carbon allocation to its biosynthesis [23]. Indeed, hyperaccumulation of Phe in *PAL*-RNAi petunia petals and *Arabidopsis* double *pal1 pal2* and quadruple *pal1 pal2 pal3 pal4* knockout mutants reduced carbon flux in the shikimate pathway, which provides precursors for Phe biosynthesis [23]. In fact, without feedback control of plastidial Phe biosynthesis by cytosolic Phe, no model is able to explain the observed accumulation of unlabeled plastidial Phe upon feeding, especially at 3 mM fed Phe (Figure 3.2) and reduction of flux through ADT, which has a low K_i value for Phe (0.10 nmol g FW⁻¹, Table 3.2). Indeed, a drastic change of cytosolic Phe concentration was predicted for 1 mM, and 3 mM fed Phe, while changes in plastidial Phe concentrations were relatively small (Table 3.1). The cytosolic Phe concentration (220 nmol g FW⁻¹, Table 3.1) was far above the predicted PCAT K_i value for Phe (34.5 nmol gFW⁻¹, Table 3.2), suggesting that during hyper-accumulation of cytosolic Phe, the plastidial Phe transporter will be feedback inhibited.

An information-theoretic approach was applied to select the model with the best performance, which suggested vacuole storage for both Phe and *p*-coumarate. Although initial pools of Phe and *p*-coumarate inside the vacuole before feeding are not identifiable through current approach, these storage mechanisms seem to function upon the feeding of exogenous Phe. It is thus hypothesized that active vacuolar transporters help maintain the flux homeostasis by translocating excess amounts of substrate/intermediates into a physically separate compartment, making them unavailable for enzymes. The role of the vacuole in sequestration of excess cytosolic Phe was recently demonstrated by [23]. Phe amount inside the vacuole was nearly doubled in *PAL*-RNAi petunia relative to wild-type control likely due the block of cytosolic Phe consumption.

Despite the low fluxes into the vacuole relative to the main flux towards lignin (Figure 3.9), the optimized model predicts that substantial accumulation of Phe and *p*-coumarate occurs in the

vacuole (Table 3.1). Indeed, the estimated maximal capacities for both vacuolar transporters are smaller (0.07 and $3.0 \text{ nmol g FW}^{-1} \text{ min}^{-1}$) than the other V_{max} values within the pathway (Table 3.2). The low capacity of Phe transporter was insufficient to efficiently translocate cytosolic Phe under high Phe (1 and 3 mM) feeding, which is the main cause of hyperaccumulation of cytosolic Phe. The most similar homologues of petunia PhCAT2 in Arabidopsis are *AtCAT2* and *AtCAT4* in stems, encoding vacuolar amino acids transporters [49]. Under normal growth conditions, they have relatively low expressions [49, 51]. However, under abiotic and biotic stresses which are known to result elevated free Phe pools in plants, the expression of *AtCAT2* and *AtCAT4* are induced [51, 52], suggesting these transporters mediate the transfer of excess Phe into vacuole.

The control coefficients of ADT, PCAT, PAL, C4H and 4CL were calculated and the results suggest PCAT has the most significant control on flux towards lignin under non-fed conditions. It has been shown previously that perturbing the expression of the plastid transporter leads to significant changes in the downstream efflux in petunia flowers [12]. The reason for this translocation step to be limiting is its relatively low predicted K_m value compared to the plastidial Phe amount (5.0 versus $15.4 \text{ nmol g FW}^{-1}$, Table 3.1 and Table 3.2). No *in vitro* enzyme assay has been done to characterize the K_m value for PCAT in Arabidopsis, although a homolog in *Escherichia coli* has a K_m value of $2 \text{ }\mu\text{M}$ [53], corresponding to around $0.4 \text{ nmol g FW}^{-1}$ based on the estimated plastid volume in plants [54]. Both calculations suggest that this transporter is easily saturated under normal condition, so that the translocation rate is largely determined by the protein amount instead of substrate level. For the other enzymatic steps, the model suggests that they are controlled by substrate levels instead of protein amounts, which explains why pathway flux is insensitive to small changes of enzyme levels in MCA. However, when a large perturbation towards the enzyme levels occurs, eventually the control will be shifted from substrate to enzyme.

Cinnamate is involved in the phenylpropanoid metabolism as a product of PAL and is the substrate for *p*-coumarate synthesis. During the feeding studies, especially at 3 mM treatment, both Phe and *p*-coumarate accumulate to a much higher amount than under the non-fed condition, suggesting that $[^{13}\text{C}_6]$ -Phe has been effectively taken up by the plant, and the pathway capacity has reached a saturation point in converting those metabolites to lignin. Even under this condition, no cinnamate was detected in extracts. Inclusion of cinnamate mass balance into the model did not improve the model's performance, while the model was able to fit the data well by simulating PAL and C4H as one single reaction. These observations might be explained by a very efficient (high

k_{cat}/K_m) C4H enzyme within the pathway. As a membrane bound protein, it is very difficult to directly test this hypothesis by measuring its maximal capacity in the basal stem tissue. Metabolic channeling has been proposed between PAL and C4H [55, 56], which provides another possible explanation of undetectable cinnamate, since the intermediate could then be efficiently passed between enzymes without having any significant accumulation.

We also attempted including mass balances for CoA and *p*-coumaroyl CoA, but found the model fits unimproved. Although the current refined model still has some underfitting issues over *p*-coumarate profiles for the 3mM treatment, it was not resolved by modeling 4CL step as a bi-substrate reaction. Another factor that might have significant effect but was not considered is ATP, which is also required by the 4CL reaction. As ATP has multiple subcellular locations, a metabolomics technology with the ability to estimate its subcellular concentrations is required.

In summary, kinetic modeling combined with a statistical evaluation procedure successfully described metabolite distributions and fluxes within the phenylpropanoid network in *Arabidopsis* lignifying cells. The AIC model selection procedure supported by feeding treatment datasets proposed regulatory mechanisms active *in vivo*, which could be experimentally tested in future. In addition, MCA suggested limitations within the phenylpropanoid network, thus providing useful guide for future metabolic engineering efforts. For example, knocking down phenylalanine export from the plastids is a first metabolic engineering target for reducing lignin biosynthesis.

3.6 References

1. Saini JK, Saini R, Tewari L: Lignocellulosic agriculture wastes as biomass feedstocks for second-generation bioethanol production: concepts and recent developments. *3 Biotech* 2015, 5(4):337-353.
2. Kim S, Dale BE: Global potential bioethanol production from wasted crops and crop residues. *Biomass and Bioenergy* 2004, 26(4):361-375.
3. Zayed H, Sahu JN, Boyce AN, Faruq G: Fuel ethanol production from lignocellulosic biomass: an overview on feedstocks and technological approaches. *Renew and Sust Energy Rev* 2016, 66:751-774.
4. Álvarez C, Reyes-Sosa FM, Díez B. Enzymatic hydrolysis of biomass from wood. *Microb Biotechnol* 2016, 9(2):149-56.

5. Pareek N, Gillgren T, Jönsson LJ: Adsorption of proteins involved in hydrolysis of lignocellulose on lignins and hemicelluloses. *Bioresour Technol* 2013, 148:70-7.
6. Li M, Pu Y, Ragauskas AJ: Current understanding of the correlation of lignin structure with biomass recalcitrance. *Front Chem* 2016, 4:45.
7. Van Acker R, Vanholme R, Storme V, Mortimer JC, Dupree P, Boerjan W: Lignin biosynthesis perturbations affect secondary cell wall composition and saccharification yield in *Arabidopsis thaliana*. *Biotechnol Biofuels* 2013, 6.
8. Wang P, Dudareva N, Morgan JA, Chapple C: Genetic manipulation of lignocellulosic biomass for bioenergy. *Curr Opin Chem Biol* 2015, 29:32-39.
9. Maeda H, Shasany AK, Schnepf J, Orlova I, Taguchi G, Cooper BR, Rhodes D, Pichersky E, Dudareva N: RNAi suppression of aroenate dehydratase1 reveals that phenylalanine is synthesized predominantly via the aroenate pathway in petunia petals. *Plant Cell* 2010, 22(3):832-849.
10. Corea OR, Ki C, Cardenas CL, Kim SJ, Brewer SE, Patten AM, Davin LB, Lewis NG: Aroenate dehydratase isoenzymes profoundly and differentially modulate carbon flux into lignins. *J Biol Chem* 2010, 287(14):11446-11459.
11. Dal Cin V, Tieman DM, Tohge T, McQuinn R, de Vos RC, Osorio S, Schmelz EA, Taylor MG, Smits-Kroon MT, Schuurink RC, *et al*: Identification of genes in the phenylalanine metabolic pathway by ectopic expression of a MYB transcription factor in tomato fruit. *Plant Cell* 2011, 23(7):2738-53.
12. Widhalm JR, Gutensohn M, Yoo H, Adebessin F, Qian Y, Guo L, Jaini R, Lynch JH, McCoy RM, Shreve JT, *et al*: Identification of a plastidial phenylalanine exporter that influences flux distribution through the phenylalanine biosynthetic network. *Nat Commun* 2015, 6:8142.
13. Yamada T, Matsuda F, Kasai K, Fukuoka S, Kitamura K, Tozawa Y, Miyagawa H, Wakasa K: Mutation of a rice gene encoding a phenylalanine biosynthetic enzyme results in accumulation of phenylalanine and tryptophan. *Plant Cell* 2008, 20(5):1316-1329.
14. Huang T, Tohge T, Lytovchenko A, Fernie AR, Jander G: Pleiotropic physiological consequences of feedback-insensitive phenylalanine biosynthesis in *Arabidopsis thaliana*. *Plant J* 2010, 63(5):823-835.

15. Zhang X, Gou M, Liu CJ: Arabidopsis Kelch repeat F-box proteins regulate phenylpropanoid biosynthesis via controlling the turnover of phenylalanine ammonia-lyase. *Plant Cell* 2013, 25:4994-5010.
16. Zhang XB, Gou MY, Guo CR, Yang HJ, Liu CJ: Down-regulation of Kelch domain-containing F-box protein in Arabidopsis enhances the production of (poly)phenols and tolerance to ultraviolet radiation. *Plant Physiol* 2015, 167:337-U548.
17. Zhang XB, Liu CJ: Multifaceted regulations of gateway enzyme phenylalanine ammonia-lyase in the biosynthesis of phenylpropanoids. *Mol Plant* 2015, 8:17-27.
18. Chen MJ, Vijaykumar V, Lu BW, Xia B, Li N: Cis- and trans-cinnamic acids have different effects on the catalytic properties of Arabidopsis phenylalanine ammonia lyases PAL1, PAL2, and PAL4. *J Integr Plant Biol* 2005, 47(1):67-75.
19. Colón AM, Sengupta N, Rhodes D, Dudareva N, Morgan J. A kinetic model describes metabolic response to perturbations and distribution of flux control in the benzenoid network of *Petunia hybrida*. *Plant J.* 2010, 62(1):64-76.
20. Hu GS, Jia JM, Hur YJ, Chung YS, Lee JH, Yun DJ, Chung WS, Yi GH, Kim TH, Kim DH: Molecular characterization of phenylalanine ammonia lyase gene from *Cistanche deserticola*. *Mol Biol Rep* 2011, 38(6):3741-50.
21. Benkeblia N, Shinano T, Osaki M: Metabolite profiling and assessment of metabolome compartmentation of soybean leaves using non-aqueous fractionation and GC-MS analysis. *Metabolomics* 2007, 3(3):297-305.
22. Krueger S, Giavalisco P, Krall L, Steinhauser MC, Büssis D, Usadel B, Flügge UI, Fernie AR, Willmitzer L, Steinhauser D: A topological map of the compartmentalized Arabidopsis thaliana leaf metabolome. *PLoS One* 2011, 6(3):e17806.
23. Lynch JH, Orlova I, Zhao C, Guo L, Jaini R, Maeda H, Akhtar T, Cruz-Lebron J, Rhodes D, Morgan J, *et al*: Multifaceted plant responses to circumvent Phe hyperaccumulation by downregulation of flux through the shikimate pathway and by vacuolar Phe sequestration. *Plant J* 2017, 92(5):939-950.
24. Zhu XG, de Sturler E, Long SP: Optimizing the distribution of resources between enzymes of carbon metabolism can dramatically increase photosynthetic rate: a numerical simulation using an evolutionary algorithm. *Plant Physiol* 2007. 145(2), 513-526.

25. Schallau K, Junker BH: Simulating plant metabolic pathways with enzyme-kinetic models. *Plant Physiol* 2010, 152(4):1763-1771.
26. Rohwer JM: Kinetic modelling of plant metabolic pathways. *J Experimental Bot* 2012, 63(6):2275-2292.
27. Beauvoit BP, Colombié S, Monier A, Andrieu MH, Biais B, Bénard C, Chéniclet C, Dieuaide-Noubhani M, Nazaret C, Mazat JP, Gibon Y: Model-assisted analysis of sugar metabolism throughout tomato fruit development reveals enzyme and carrier properties in relation to vacuole expansion. *Plant Cell* 2014, 26(8):3224-42.
28. Curien G, Bastien O, Robert-Genthon M, Cornish-Bowden A, Cárdenas ML, Dumas R: Understanding the regulation of aspartate metabolism using a model based on measured kinetic parameters. *Mol Syst Biol* 2009, 5:271.
29. Fell DA: Metabolic control analysis. *Syst Biol* 2005, 13(5):69-80.
30. Haunschild MD, Freisleben B, Takors R, Wiechert W: Investigating the dynamic behavior of biochemical networks using model families. *Bioinformatics* 2005, 21(8):1617-1625.
31. Akaike H: A new look at the statistical model identification. *IEEE Trans Autom Control* 1974, 19(6):716-723.
32. Burnham KP, Anderson DR: Model selection and multi-model inference. a practical informatio-theoric approach. *Springer* 2003, 1229.
33. Link H, Kochanowski K, Sauer U: Systematic identification of allosteric protein-metabolite interactions that control enzyme activity in vivo. *Nat Biotechnol*, 2013, 31(4):357-361.
34. Lee Y, Escamilla-Treviño L, Dixon RA, Voit, EO: Functional analysis of metabolic channeling and regulation in lignin biosynthesis: a computational approach. *PLoS Comput Biol*, 2012, 8(11).
35. Wang JP, Naik PP, Chen HC, Shi R, Lin CY, Liu J, Shuford CM, Li Q, Sun YH, Tunlaya-Anukit S, *et al*: Complete proteomic-based enzyme reaction and inhibition kinetics reveal how monolignol biosynthetic enzyme families affect metabolic flux and lignin in *Populus trichocarpa*. *Plant Cell* 2014, 26(3):894-914.
36. Faraji M, Fonseca LL, Escamilla-Treviño L, Dixon, RA, Voit EO: Computational inference of the structure and regulation of the lignin pathway in *Panicum virgatum*. *Biotechnol Biofuels* 2015, 8(1):151.

37. Wang P, Guo L, Jaini R, Klempien A, McCoy RM, Morgan JA, Dudareva N, Chapple C: A ^{13}C isotope labeling method for the measurement of lignin metabolic flux in *Arabidopsis* stems. *Plant Methods* 2018, 14:51.
38. Jaini R, Wang P, Dudareva N, Chapple C, Morgan JA: Targeted metabolomics of the phenylpropanoid pathway in *Arabidopsis thaliana* using reversed phase liquid chromatography coupled with tandem mass spectrometry. *Phytochem Anal* 2017, 28(4):267-276.
39. Lu F, Ralph J: Derivatization followed by reductive cleavage (DFRC method), a new method for lignin analysis: protocol for analysis of DFRC monomers. *J Agric Food Chem* 1997, 45(7):2590-2592.
40. Chang X, Chandra R, Berleth T, Beatson RP: Rapid, microscale, acetyl bromide-based method for high-throughput determination of lignin content in *Arabidopsis thaliana*. *J Agric Food Chem* 2005, 56(16):6825-6834.
41. Haario H, Laine M, Mira A, Saksman E: DRAM: efficient adaptive MCMC. *Stat Comput* 2006, 16(4), 339–354.
42. Bonawitz ND, Chapple C: The genetics of lignin biosynthesis: connecting genotype to phenotype. *Annu Rev Genet* 2010, 44(1):337-363.
43. Lee D, Meyer K, Chapple C, Douglas CJ: Antisense suppression of 4-coumarate: coenzyme A ligase activity in *arabidopsis* leads to altered lignin subunit composition. *Plant Cell* 1997, 9(11):1985-1998.
44. Ehlting J, Büttner D, Wang Q, Douglas CJ, Somssich IE, Kombrink E: Three 4-coumarate: coenzyme A ligases in *Arabidopsis thaliana* represent two evolutionarily divergent classes in angiosperms. *Plant J* 1999, 19(1):9-20.
45. Rohde A, Morreel K, Ralph J, Goeminne G, Hostyn V, De Rycke R, Kushnir S, Van Doorselaere J, Joseleau JP, Vuylsteke M *et al*: Molecular phenotyping of the *pall* and *pal2* mutants of *Arabidopsis thaliana* reveals far-reaching consequences on phenylpropanoid, amino acid, and carbohydrate metabolism. *Plant Cell* 2004, 16(10):2749-2771.
46. Cochrane FC, Davin LB, Lewis NG: The *Arabidopsis* phenylalanine ammonia lyase gene family: kinetic characterization of the four PAL Isoforms. *Phytochemistry* 2004, 65(11):1557-1564.

47. Li Y, Kim JI, Pysh L, Chapple C: Four isoforms of *Arabidopsis thaliana* 4-coumarate: CoA ligase (4CL) have overlapping yet distinct roles in phenylpropanoid metabolism. *Plant Physiol* 2015, 4(765):2409-2421.
48. Su Y, Frommer WB, Ludewig U: Molecular and functional characterization of a family of amino acid transporters from *Arabidopsis*. *Plant Physiol* 2004, 136(2):3104-3113.
49. Yang H, Krebs M, Stierhof Y, Ludewig U: Characterization of the putative amino acid transporter genes *AtCAT2*, 3 & 4: The tonoplast localized *AtCAT2* regulates soluble leaf amino acids. *J Plant Physiol* 2014, 171(8):594-601.
50. Heinig U, Gutensohn M, Dudareva N, Aharoni A: The challenges of cellular compartmentalization in plant metabolic engineering. *Curr Opin Biotechnol* 2013, 24(2):239-246.
51. Winter D, Vinegar B, Nahal H, Ammar R, Wilson GV, Provart NJ: An “electronic fluorescent pictograph” browser for exploring and analyzing large-scale biological data sets. *PloS One* 2007, 2(8):e718.
52. Návarová H, Bernsdorff F, Döring A, Zeier J: Pipecolic acid, an endogenous mediator of defense amplification and priming, is a critical regulator of inducible plant immunity. *Plant Cell* 2012, 24(12):5123-5141.
53. Pi J, Wookey PJ, Pittard AJ: Cloning and sequencing of the *pheP* gene, which encodes the phenylalanine-specific transport system of *Escherichia coli*. *J Bacteriol* 1991, 173(12):3622-3629.
54. Winter, H., Robinson, D. G., & Heldt, H. W. (1993). Subcellular Volume and Metabolic Concentrations in Barley Leaves. *Planta* 1993, 191:180-190.
55. Rasmussen S, Dixon RA: Transgene-mediated and elicitor-induced perturbation of metabolic channeling at the entry point into the phenylpropanoid pathway. *Plant Cell*, 1999, 11(8):1537-1552.
56. Achnine L, Blancaflor EB, Rasmussen S, Dixon RA: Colocalization of L-phenylalanine ammonia-lyase and cinnamate 4-hydroxylase for metabolic channeling in phenylpropanoid biosynthesis. *Plant Cell* 2004, 16(11):3098-3109.

Table 3.1 Subcellular metabolite profiles under different feeding conditions predicted by the refined mode

(Units: nmol g FW⁻¹, values are represented as average \pm standard deviation estimated by MCMC)

Condition	t=0 min		t=240 min		
Metabolite (Location)	Non-fed	0.1 mM	0.3 mM	1 mM	3 mM
Phe (Xylem)	0.0 \pm 0.0	2.8 \pm 0.3	8.4 \pm 0.8	27.8 \pm 2.6	83.5 \pm 7.7
Phe (Plastid)	15.4 \pm 0.9	15.4 \pm 0.8	15.5 \pm 0.8	16.1 \pm 0.8	25.0 \pm 2.1
Phe (Cytosol)	1.7 \pm 0.6	2.3 \pm 0.8	4.0 \pm 1.3	12.6 \pm 3.9	220 \pm 21
Phe (Vacuole)	0.00 \pm 0.00	9.4 \pm 1.0	11.2 \pm 1.2	14.3 \pm 2.7	16.4 \pm 4.3
<i>p</i> -Coumarate (Cytosol)	0.07 \pm 0.01	0.10 \pm 0.01	0.16 \pm 0.02	0.40 \pm 0.04	1.11 \pm 0.13
<i>p</i> -Coumarate (Vacuole)	0.00 \pm 0.00	0.93 \pm 0.04	1.5 \pm 0.1	3.7 \pm 0.2	9.3 \pm 0.7

Table 3.2 Final parameter list estimated by the refined model

Parameter Identity	Reference value	Estimated value	95% Confidence	Unit
Total Phe amount at t=0	16.7±3.9 [#]	17.2	[16.3, 17.8]	nmol g FW ⁻¹
Total pCA amount at t=0	0.3±0.2 [#]	0.07	[0.06, 0.09]	nmol g FW ⁻¹
Cytosolic fraction of Phe at t=0	-	0.06	[0.04, 0.18]	dimensionless
Xylem volume	-	31.7	[22.6, 32.6]	μL g FW ⁻¹
$K_{M,Phe}^{PCAT}$	-	5.0	[0.4, 9.8]	nmol g FW ⁻¹
$K_{M,Phe}^{PXT}$	-	56.3	[45.6, 76.9]	nmol g FW ⁻¹
$K_{M,Phe}^{PAL}$	4.3 [*]	3.3	[2.2, 12.0]	nmol g FW ⁻¹
$K_{M,Phe}^{VCAT}$	-	1.2	[0.17, 4.3]	nmol g FW ⁻¹
$K_{M,pCA}^{4CL}$	2.4 ^{**}	0.30	[0.27, 0.48]	nmol g FW ⁻¹
$K_{M,pCA}^{PVT}$	-	70.0	[20.7, 98.9]	nmol g FW ⁻¹
$K_{i,Phe}^{ADT}$	-	0.10	[0.10, 0.73]	nmol g FW ⁻¹
$K_{i,Phe}^{PCAT}$	-	34.5	[5.2, 94.8]	nmol g FW ⁻¹
$V_{max,ADT}$	-	548	[75.1, 537]	nmol g FW ⁻¹ min ⁻¹
$V_{max,PCAT}$	-	4.6	[3.6, 5.7]	nmol g FW ⁻¹ min ⁻¹
$V_{max,PXT}$	-	20.4	[18.9, 28.6]	nmol g FW ⁻¹ min ⁻¹
$V_{max,PAL}$	10.5±3.0 [#]	14.7	[14.3, 17.5]	nmol g FW ⁻¹ min ⁻¹
$V_{max,VCAT}$	-	0.07	[0.04, 0.11]	nmol g FW ⁻¹ min ⁻¹
$V_{max,4CL}$	82.4±21.4 [#]	19.0	[18.0, 23.4]	nmol g FW ⁻¹ min ⁻¹
$V_{max,PVT}$	-	3.0	[0.74, 4.2]	nmol g FW ⁻¹ min ⁻¹

[†]Estimated with 100,000 MCMC samples

[#]Measured in current study

^{*}Cochrane et al. (2004), cytosol volume estimated from Winter et al. (1993)

^{**}Ehlting et al. (1999)

Table 3.3 AIC scores of the models relative to the one without any potential metabolite-enzyme interactions.

Shown are the relative change of AIC scores when each interaction was introduced into the refined model. No interaction significantly improves the model's performance since all AIC score changes are positive.

Introduced Interaction	Δ AIC
Phev competitively inhibits VCAT	0.32
Phec competitively inhibits PXT	0.33
Phev uncompetitively inhibits VCAT	0.46
<i>p</i> CAC uncompetitively inhibits PXT	1.57
Phec uncompetitively inhibits PCAT	2.25
<i>p</i> CAC uncompetitively inhibits PAL	2.25
<i>p</i> CAV uncompetitively inhibits PVT	2.59
<i>p</i> CAC competitively inhibits PAL	2.68
<i>p</i> CAC uncompetitively inhibits PCAT	2.69
<i>p</i> CAV competitively inhibits PVT	2.74
<i>p</i> CAC competitively inhibits PXT	2.74
<i>p</i> CAC competitively inhibits PCAT	2.76
Phec activates 4CL	2.79
Phec uncompetitively inhibits PXT	18.04

Table 3.4 Flux control coefficients of pathway enzymes

Enzyme	Flux Control Coefficient	95% Confidence intervals [†]
ADT	0.19	[0.03, 0.28]
PCAT	0.80	[0.71, 0.97]
PAL	0.01	[0.01, 0.02]
4CL	0.00	[0.00, 0.00]
VCAT	-0.01	[-0.01, -0.01]
PVT	0.00	[0.00, 0.00]

[†]Estimated with 100,000 MCMC samples

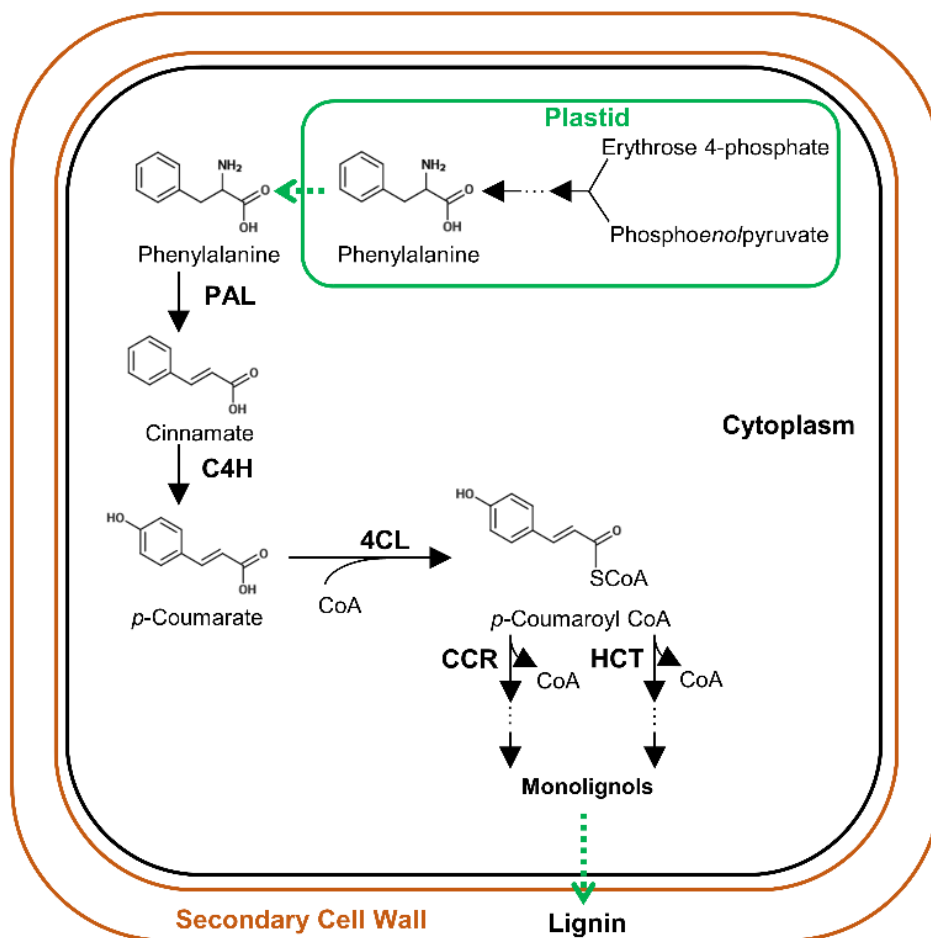


Figure 3.1 Lignin biosynthesis in Arabidopsis.

A simplified schematic of lignin biosynthesis is shown with metabolic steps explicitly simulated in the current study. Black arrows represent metabolic reactions with corresponding enzyme names. Green dashed arrows represent membrane transport steps between cellular compartments.

Abbreviations: 4CL, 4-coumarate: CoA ligase; C4H, cinnamate 4-hydroxylase; CCR, cinnamoyl-CoA reductase; HCT, hydroxycinnamoyl CoA: shikimate hydroxycinnamoyl transferase; PAL, phenylalanine ammonia lyase.

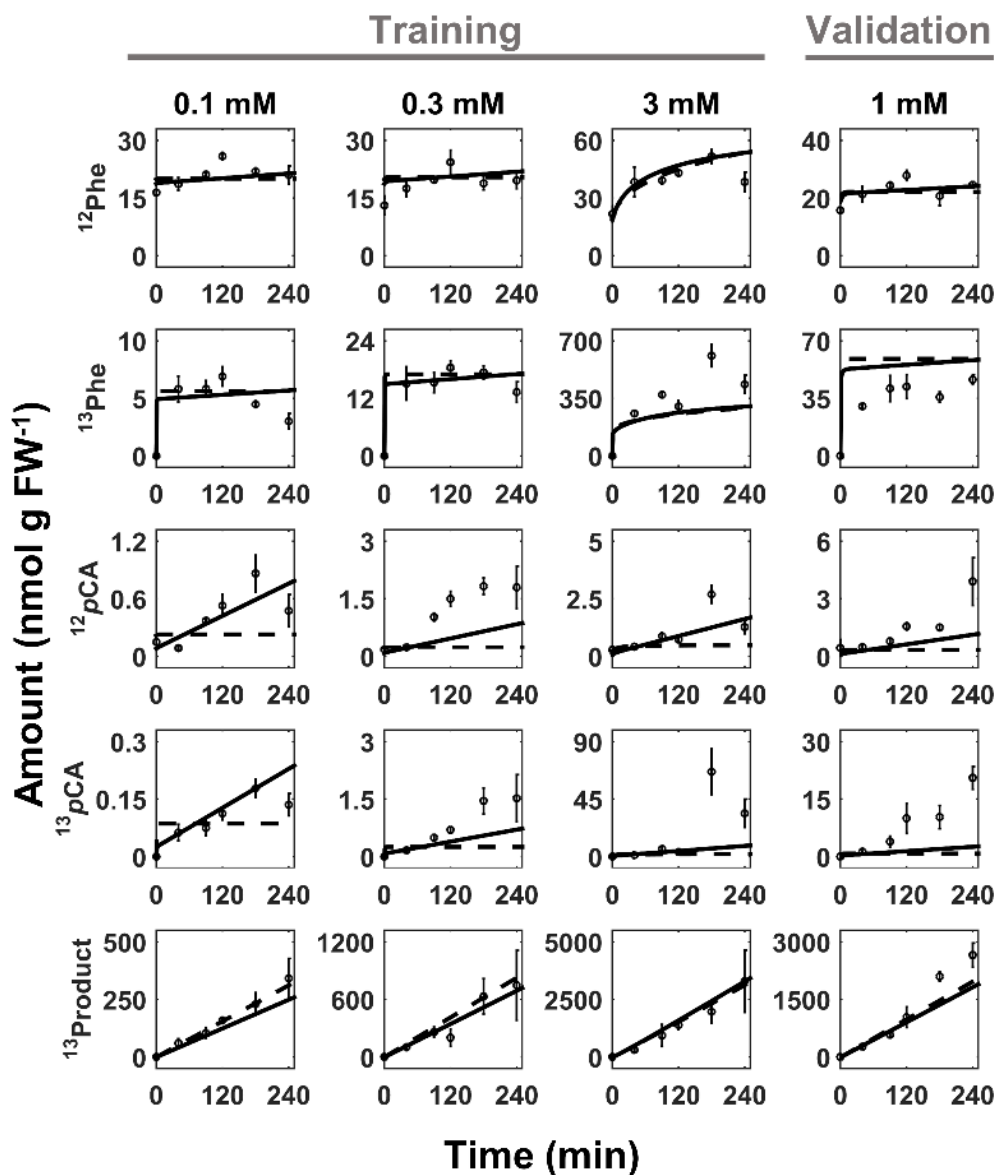


Figure 3.2 Data fitting comparison between base and refined models.

The model performance with experimental observations are shown for both base model (dashed) and refined model (solid) with vacuole storage for both phenylalanine and *p*-coumarate. Three treatments were used for both model training, with the 1 mM dataset left out for validation.

Superscript 12 represents natural molecules with ^{12}C , while 13 represents molecules with ^{13}C labeled rings. $^{13}\text{Product}$ is the sum of all quantified phenylpropanoid molecules including lignin monomers labeled with ^{13}C . Details of experiments can be found in Materials and Methods.

Measurements are the average \pm standard deviation ($n=3$), while model predictions are shown as lines. Abbreviations: *pCA*, *p*-coumarate.

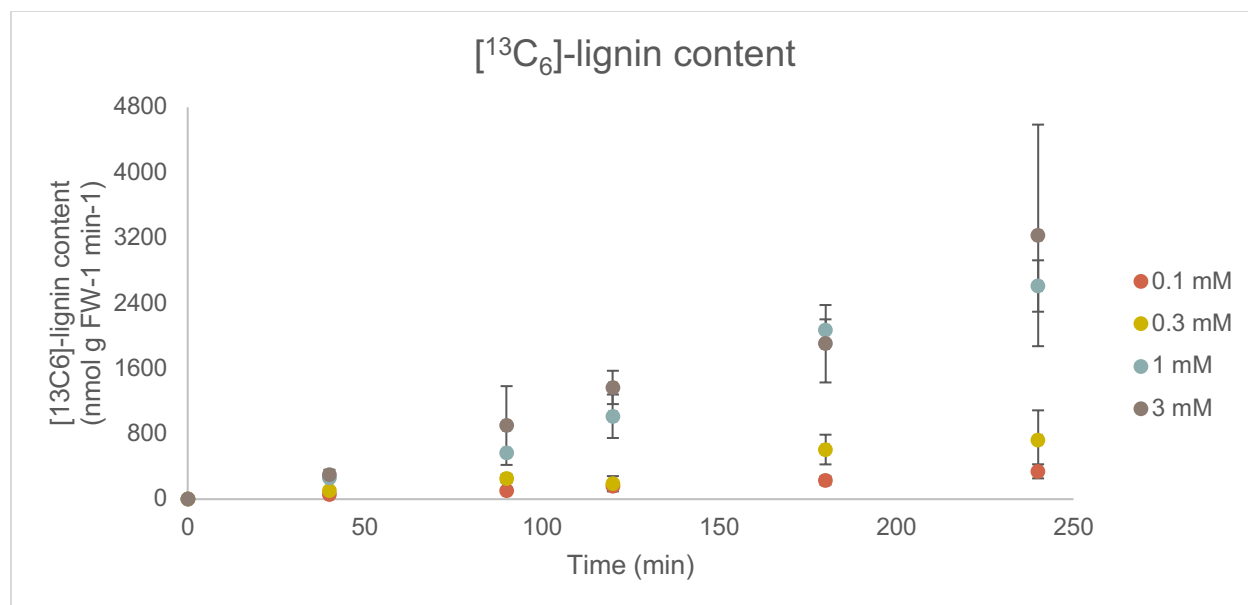


Figure 3.3 Total [¹³C₆]-lignin in stems fed with [¹³C₆]-Phe.

Measurements are represented as points by average \pm standard deviation (n=3).

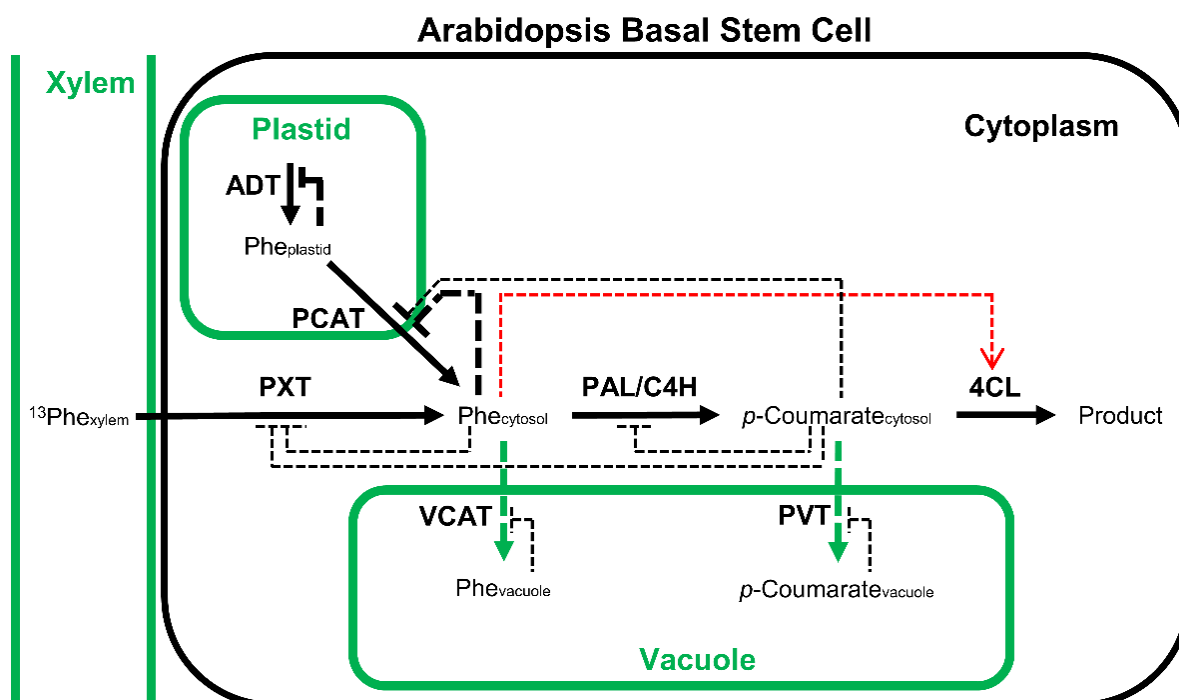


Figure 3.4 Metabolite-enzyme interactions and vacuole storage mechanisms considered in the general phenylpropanoid pathway.

The kinetic model in the current study is depicted with possible interactions between each metabolite and enzyme, and two possible vacuole storage mechanisms. Red dashed arrow represents feedforward activation, black dashed lines represent both competitive and uncompetitive feedback inhibition, green dashed arrows represent vacuole storage fluxes.

Feedback competitive inhibition of plastidial phenylalanine towards ADT and cytosolic phenylalanine towards PCAT are bolded as they are included in the model *a priori*.

Abbreviations: PCAT, plastidial cationic amino-acid transporter; PVT, putative vacuolar transporter; PXT, putative xylem transporter; VCAT, vacuolar cationic amino-acid transporter.

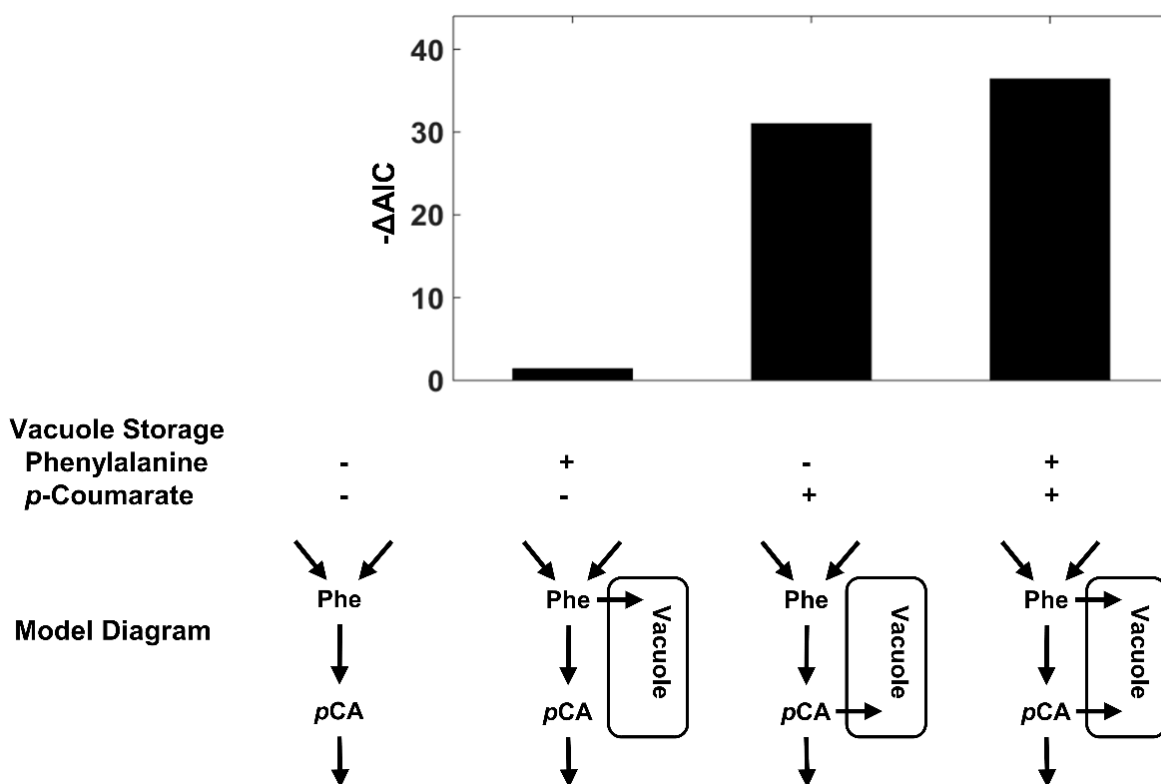


Figure 3.5 Reduction in the AIC score relative to the base model without vacuole storage.

Three additional models were generated and tested by considering possible combinations of vacuole storage mechanisms within the studied metabolism. Corrected AIC was applied to evaluate the performance of these models. The model with lowest AIC score is the one which fits the data best with the simplest structure.

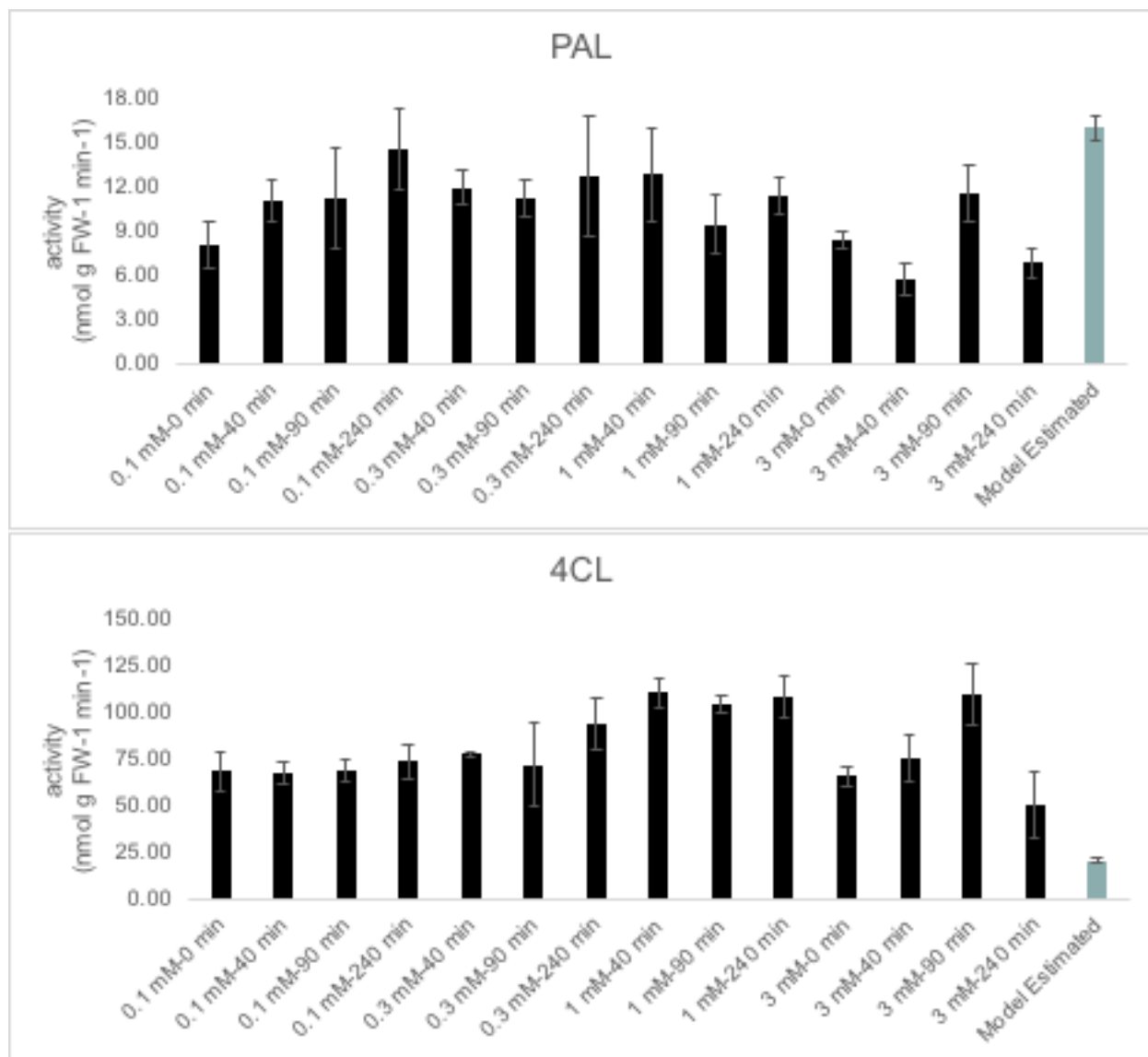


Figure 3.6 Measured and estimated PAL and 4CL activities in wild type stems over the feeding period.

Shown are the maximal capacities of PAL and 4CL, which were obtained from measurements within the feeding period (black bars), or refined model prediction (grey bars). All bars are presented as average \pm standard deviation ($n=3$ for measurements).

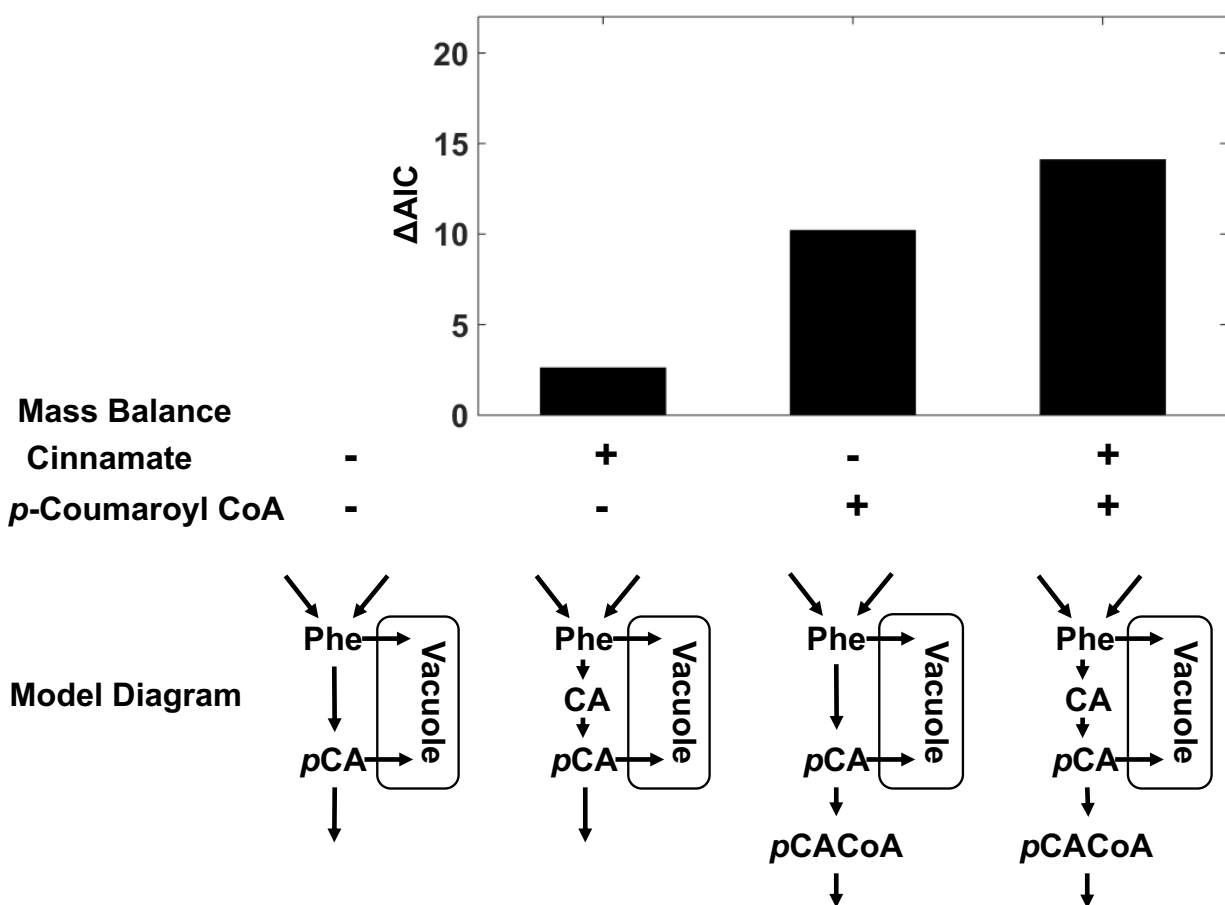


Figure 3.7 AIC scores of the models relative to one without cinnamate and *p*-coumaroyl CoA mass balances.

Three additional models were generated by considering all possible combinations of mass balances of cinnamate and *p*-coumaroyl CoA. Corrected AIC was applied to evaluate the performance of these models.

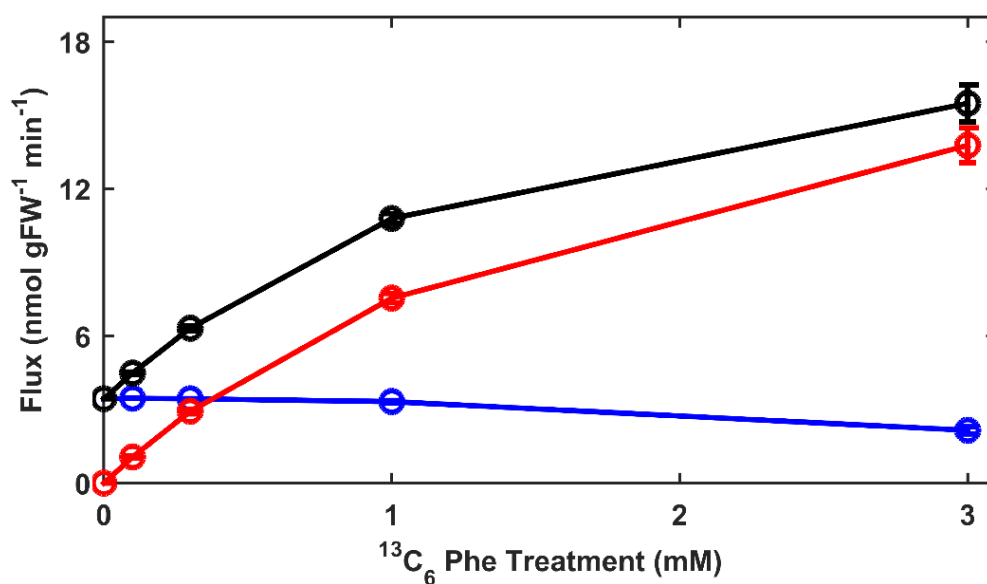


Figure 3.8 Pathway flux distribution at 240 min for different feeding treatments.

Fluxes at 240 min for different feeding treatments were predicted by the refined model. Shown were the average \pm standard deviation values for fluxes from 100,000 MCMC simulation samples. Black line represents total carbon flux through the 4CL enzyme (sum of ^{12}C and ^{13}C fluxes), red line is the flux through the PXT, and the blue line is the reaction rate of ADT.

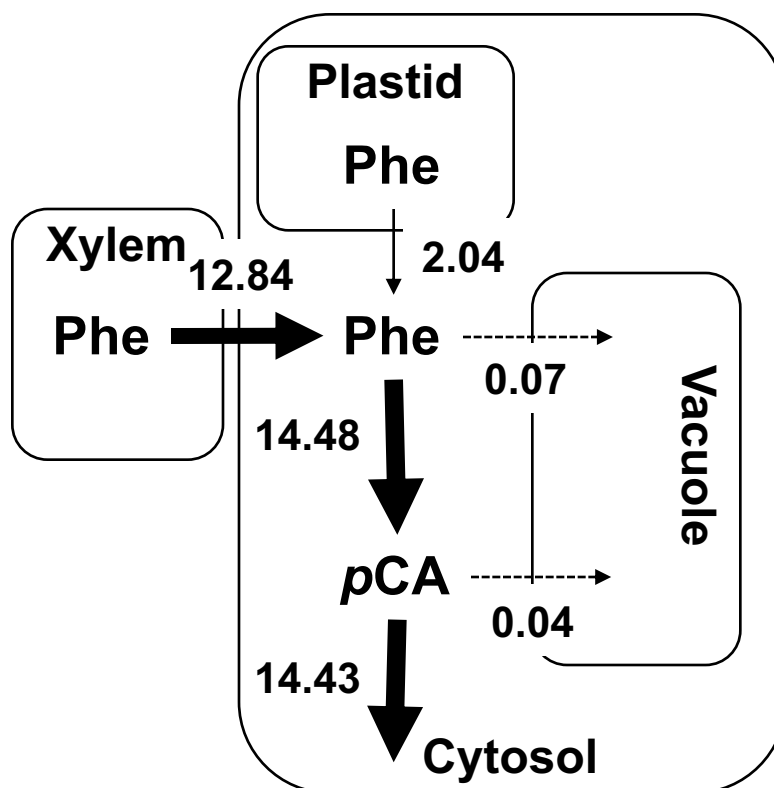


Figure 3.9 Flux distribution of the phenylpropanoid pathway during 3 mM $^{13}\text{C}_6$ Phe feeding condition.

Shown were the fluxes at 240 min after feeding, predicted by the refined model. The width of the arrows is proportional to the flux through the reactions (Units: $\text{nmol g FW}^{-1} \text{ min}^{-1}$).

CHAPTER 4. THE TRANSCRIPTS AND METABOLITES NETWORK IN RESPONSE TO LIGNIN BIOSYNTHESIS PERTURBATIONS IN ARABIDOPSIS

4.1 Introduction

Lignin, a phenolic polymer intertwined with polysaccharides in the secondary cell wall, equips plant cells with mechanical strength and the hydrophobicity required to conduct water. In angiosperms, lignin is constituted of three monomers, namely *p*-coumaryl, coniferyl, and sinapyl alcohol, which form *p*-hydroxyphenyl (H), guaiacyl (G), and syringyl (S) lignin, respectively [1]. These monomers are derived from the phenylpropanoid pathway starting with phenylalanine (Phe). In monocots, tyrosine (Tyr) is an additional precursor [2]. The side chain of the amino acids is deaminated and reduced, and the phenyl ring is hydroxylated and *O*-methylated, to give rise to the three monolignols through eleven enzymatic steps. These enzymes include Phe ammonia lyase (PAL) or PAL/Tyr ammonia lyase in monocots, cinnamate 4-hydroxylase (C4H), 4-coumarate:CoA ligase (4CL), hydroxycinnamoyl:CoA shikimate hydroxycinnamoyl transferase (HCT), *p*-coumaroyl 3'-hydroxylase (C3'H), caffeoyl shikimate esterase (CSE), caffeoyl CoA *O*-methyltransferase (CCoAOMT), ferulate 5-hydroxylase (F5H), caffeic acid *O*-methyltransferase (COMT), cinnamoyl CoA reductase (CCR), and cinnamyl alcohol dehydrogenase (CAD). The monomers, once synthesized, are exported into the cell wall and polymerized into lignin. Polymerization of monolignols requires laccases (LAC) and peroxidases (PRX) [3, 4].

Phenylpropanoid metabolism is tightly regulated at transcriptional and post-translational levels to control the significant carbon flux towards lignin. A hierarchical network of transcription factors (TFs) regulates lignification and secondary cell wall formation in plants. Five master regulators from the NAM, ATAF1,2, and CUC2 (NAC) family activates the secondary-level MYB TFs including MYB46 and MYB83 [5, 6]. These master switches together turn on the downstream TFs to activate (with MYB58, MYB63, MYB85) or repress (with MYB4) lignification [7]. In addition to the TFs, the Mediator complex maintains phenylpropanoid homeostasis via bridging the TFs and the transcriptional machinery. Disruption of Mediator subunits *MED5a* and *MED5b* (*MED5a/b*) in *Arabidopsis* causes increased levels of phenylpropanoids and transcripts of the biosynthetic genes involved [8, 9]. Post-translationally, four kelch-repeat-containing F-box (KFB) proteins mediate the proteolytic degradation of PAL enzymes in plants, controlling the entry step

for phenylpropanoid production [10, 11]. Similarly, another KFB protein (KFB^{CHS}) targets chalcone synthase (CHS, catalyzing the first committed step for flavonoids production) to regulate the biosynthesis of anthocyanins and flavonols [12].

Alterations in transcript abundance of lignin biosynthetic genes lead to changes in enzyme activities and subsequent metabolite accumulation [13, 14]. Changed phenylpropanoid content, in turn, can result in differential gene expression in plants. It has been shown that *trans*-cinnamate, the product of PAL, inhibits the transcription of three *PAL* genes in bean (*Phaseolus vulgaris* L.) [15]. *cis*-cinnamate, readily converted from *trans*-cinnamate dependent on UV, causes auxin-like phenotype and triggers induction of early auxin responsive genes in Arabidopsis seedlings [16]. Abolition of *MED5a/b* expression in *C3'H* deficient mutants rescues their dwarf phenotype, an observation that has led to a model in which altered metabolites levels are sensed directly or indirectly by Mediator to influence gene expression [9].

Genetic modification of the genes involved in lignification often leads to altered lignin content and/or composition [17, 18]. Numerous studies of natural mutants or genetically manipulated plants have demonstrated that perturbations at each step of lignification result in global transcription reprogramming and shifts in primary and secondary metabolism [9, 13, 19]. In an attempt to investigate metabolic activities of the whole lignin biosynthetic pathway, we sought to obtain a picture of the global transcriptional alterations in response to defects in lignification. The transcriptome data, combined together with metabolome information, can illuminate the impacts of pathway perturbations on carbon flux towards and within the phenylpropanoid network. The rich collection of mutants makes Arabidopsis a good model in which to systematically study how blockage in particular steps of lignification influences gene expression and carbon re-allocation, like the systems biology approach previously employed linking microarray and metabolomic data from plants defective in ten lignin biosynthetic steps [19]. Here we selected double mutants with stronger deficiency in *PAL* (*pal1 pal2*) [13], and *CAD* (*cadC cadD*) [20]; a *4CL1* mutant (*4cl1*) that is defective in the last step of the general phenylpropanoid metabolism [14]; a mutant of the recently identified lignin biosynthetic gene *CSE* (*cse-2*) [21]. We also included two allelic mutants of *C4H* (a severe mutant called *reduced epidermal fluorescence 3-2* (*ref3-2*), and a moderate mutant *ref3-3*) [22]. Three mutants were chosen with different expression of *F5H* which determines the composition of G and S lignin [23], namely a knock-out mutant (*fah1*) enriched in G lignin, an over-expresser of *F5H* gene (*C4H:F5H*)

enriched in S lignin, and a mutant that bypasses *C3'H* and expresses *F5H* from *Selaginella moellendorffii* (*ref8 fah1 SmF5H*) to generate mainly H and S lignin [23-25]. Two of the remaining mutants namely *med5a/b*, *med5a/b ref8*, were defective in the MED5 subunit of the Mediator complex and *C3'H (ref8)* [9], and the final two were defective in *ref2 (cyp83a1)* a mutant of glucosinolate biosynthesis that accumulates reduced levels of phenylpropanoids as well as *med5a/b ref2* a mutant combination that suppresses *ref2* [26].

We conducted transcriptome analysis using RNA sequencing (RNA-seq) in these selected mutants with the wild-type controls and surveyed their phenylpropanoid profiles. We show that the 13 perturbed mutant lines all exhibited transcriptional and metabolic responses to the genetic mutations, although most of them were morphologically indistinguishable from wild type. Defects in lignin biosynthesis triggered systematic induction of genes involved in lignification and Phe supply, however these genes were repressed in *fah1* and *cse-2*. The similarity of gene expression patterns identified five subgroups of mutants with shared as well as distinct metabolite accumulation. In addition to changes related to phenylpropanoid metabolism, these mutants exhibited transcriptional reprogramming associated with various stress responses. Combined together, the transcriptome and metabolite data identified the molecular responses to perturbed lignin biosynthesis in plants that will provide new insights to analyze flux re-distribution.

4.2 Results

4.2.1 Global transcriptome analysis in Arabidopsis with perturbed lignification

To investigate the effects of lignin perturbation on transcripts and metabolites, we selected 13 Arabidopsis mutants together with wild type to collect the data. The Arabidopsis plants were grown under the same condition and harvested at the same time when the inflorescence stems of most plants were approximately 10 cm high. We collected the basal 0.5-2 cm of the primary stems as lignin is being deposited rapidly in this tissue at this stage [27]. We harvested tissue for transcriptome analysis after 28 days of planting and tissue for metabolite quantification on the next day. At harvest, the morphology of most of the mutants was like wild type, except for *ref3-2*, *ref3-3*, and *ref8 fah1 SmF5H* (Figure 4.1). Both *ref3* mutants, defective in *C4H*, had narrow and down-curved leaves, a phenotype previously seen in high auxin mutants [28]. *ref3-3* showed similar

height compared to wild type, whereas *ref3-2* was dwarfed, as reported before [22]. Like *ref3-2*, *ref8 fah1 SmF5H* was also shorter than wild type [25].

We conducted an RNA-seq experiment to obtain transcriptome data from the stem tissue. To increase the robustness of the statistical analysis, we included 8 replicates of wild type controls and 4 replicates for each mutant. After the transcript reads were mapped to the Arabidopsis genome, the whole dataset contained 20974 expressed genes (Criteria for a gene to be considered as expressed are explained in Methods). An analysis using STRING [29] of the 50 genes with highest counts across all samples revealed that these genes were functionally involved in lignin biosynthesis, photosynthesis, glucosinolate biosynthesis, *S*-adenosylmethionine biosynthesis, and cellulose biosynthesis (Figure 4.2). Indeed, the major isoforms of lignin biosynthetic genes showed high expression among all expressed genes in wild type plants (Table 4.1). These observations agreed that the basal stem tissue sampled was undergoing active lignification.

To examine the data quality of the replicates and similarity of the expression profiles of all the samples, we performed principle component analysis (PCA) with the transcriptome data of individual samples (Figure 4.3). The first two principle components (PC) together explained about 60% variance. Interestingly, the 8 wild type replicates clustered less tightly than the mutants on PC1 and PC2. This result suggested that variance existed among wild type. The mutant samples clustered together based on genotypes, suggesting that they were distinguished with their transcriptome profiles. *pal1 pal2*, *fah1*, *4cl1*, *cse-2*, and *ref2* samples scattered around the wild-type samples (Figure 4.3), suggesting that their overall expression profiles are similar to the wild-type controls. The two *ref3* allelic mutants were close to each other and separated from the wild-type samples on PC1. An independent PCA analysis with only wild type, *ref3-2*, and *ref3-3* showed that the *ref3-3* samples, from the weaker of the two mutants, were closer to wild type on PC1 than those from *ref3-2*, the stronger mutant (Figure 4.4). This pattern suggested that the more severe defect in C4H in *ref3-2* caused greater changes in transcriptome than that in *ref3-3*. *C4H:F5H*, *cadC cadD*, *ref8 fah1 SmF5H* clustered close to one another, indicating they share similar changes in gene expression patterns (Figure 4.3). The three mutants containing *med5* mutations formed another subgroup, suggesting that alterations in genes directly or indirectly regulated by *MED5* differentiated the samples from all other genotypes. A separate PCA of only the three *med5*-containing mutants with wild type revealed that they were separated from each other on PC2 (Figure 4.5), indicating distinct expression patterns caused by the additional mutation in *REF8* or

REF2. To independently evaluate the PCA-based clustering results between individual samples and identify their similarity, we clustered samples based on the Euclidean distance from all genes (Figure 4.6). Similar to the PCA result, the 8 wild-type samples were clustered into two groups containing different mutants, again showing variance among the replicates. *pal1 pal2*, *fah1*, *4cl1*, *cse-2*, and *ref2* samples were closely clustered with wild-type samples. *ref3-3* samples were more similar to wild type than *ref3-2*, supporting the suggestion that the weak allele of *C4H* caused fewer and/or more subtle changes than the strong allele.

4.2.2 Accumulation of phenylpropanoids changed in mutants

To assess how the perturbations in lignification affect the levels of phenylpropanoids, we extracted soluble metabolites from the basal stem tissue of wild type and the 13 mutant plants and quantified 18 compounds on LC-MS/MS [30]. Comparison of each metabolite in these mutants to the wild-type controls revealed substantial changes in the accumulation of numerous phenylpropanoids (Figure 4.7, Table 4.2). *fah1* and *ref2*, similar to how they clustered with wild type in terms of their transcriptional changes, grouped together with wild type, indicating they have near normal metabolite profiles. In contrast, the *pal1 pal2* double mutant contained 660 nmol g FW⁻¹ Phe, about 25 times higher than wild type. This substantial accumulation of Phe marked a distinct feature of *pal1 pal2*, while the remaining compounds downstream of PAL showed similar or reduced levels compared to wild type, due to a decreased influx into the pathway. Similarly, hyper-accumulation of immediate substrates was observed in the other perturbed plants: the *4cl1* mutant displayed enhanced contents of *p*-coumarate and ferulate and *cse-2* was marked by elevated levels of caffeate and caffeoyl shikimate. Concentrations of coniferaldehyde and sinapaldehyde were increased in *cadC cadD*. In both *med5a/b ref8* and *ref8 fah1 SmF5H* plants *p*-coumaroyl derivatives accumulated at higher levels when *C3'H* is mutated. Although cinnamate was below the detection limit in *ref3* mutants using the current method, caffeoyl CoA, downstream of *C4H*, showed a substantial reduction in *ref3-2*. Overexpression of F5H in *C4H:F5H* led to significant increase in sinapaldehyde but no difference in sinapyl alcohol. *med5a/b* displayed increased accumulation of caffeoyl CoA, coniferyl alcohol and sinapyl alcohol, consistent with the repressive role of *MED5a/b* in phenylpropanoid metabolism.

4.2.3 Transcription of genes involved in lignification altered in mutants

To better understand the relationships between altered metabolite levels and the transcription of genes of the lignin biosynthetic pathway, we compared the expression of every gene in each mutant to wild type and extracted the \log_2 fold change (\log_2FC) for the target genes (Figure 4.8). The expression of mutated genes in the T-DNA mutants (*pal1 pal2*, *4cl1*, *cse-2*, *cadC cadD*) and *fah1* (harboring a nonsense mutation that is known to trigger nonsense-mediated decay [31]) was significantly reduced. *ref3-2*, *ref3-3*, and *ref8* mutants, isolated from an ethyl methanesulfonate mutated population, each contain a missense transition mutation [22, 32]. Mutated nucleotides in *ref3-2*, *ref3-3*, and *ref8* were revealed in the assembled transcript sequences, although the corresponding transcripts showed increased levels (Figure 4.8). Distinct from the mutated genes, transcript abundance of the other lignin biosynthetic genes was increased or not changed in these biosynthetic mutants. *cse-2* and *fah1* were two exceptions which showed decreased levels of transcripts of most of the other genes in the pathway. In *pal1 pal2*, *PAL3* was slightly up-regulated. Although it has been reported that *4CL1* was induced in *pal1 pal2* in the ecotype of C24 [13], we found that *4CL2* and *4CL3* instead exhibited higher expression in Col-0 (Figure 4.8). The *ref3* mutants showed elevated levels of transcript for most of the other lignin biosynthetic genes. Among the differentially expressed lignin biosynthetic genes, *PAL2*, *4CL2*, *4CL4*, and *CCR2* displayed more substantially altered transcription in more mutants than their respective homologs. *4CL3*, with a dominant role in flavonoid biosynthesis [14], was induced in *fah1* but repressed in *C4H:F5H*, in contrast to the other *4CL* isoforms (Figure 4.8). When *MED5a/b* was disrupted in *med5a/b*, *med5a/b ref8*, and *med5a/b ref2* phenylpropanoid biosynthetic genes showed increased expression, in line with the repressive role of *MED5a/b* in pathway homeostasis [9, 33]. Although the suppression of phenylpropanoid metabolism in *ref2* occurs primarily through the degradation of PAL enzymes (Jeong Im Kim, unpublished data), the transcript levels of all the major isoforms of lignin biosynthetic genes in *ref2* were decreased, suggesting additional transcriptional regulation of phenylpropanoid pathway in this mutant.

Disruption in a number of *LAC* and *PRX* genes involved in the polymerization of monolignols results in decreased lignin content as well as altered soluble phenylpropanoids in plants [3, 4, 34]. In our mutants, defects in lignin biosynthetic genes caused altered expression of *LAC4*, *LAC11*, *LAC17*, *PRX64*, and *PRX72* (Figure 4.8). Interestingly, in *med5a/b*, *med5a/b ref8*, *med5a/b ref2*, the three *LAC* genes and *PRX64* were down-regulated, suggesting that in wild type

plants Mediator activates their expression, distinct from the role it plays in regulating monolignol biosynthetic genes.

We next examined the genes involved in shikimate biosynthesis to determine whether the production of Phe was affected at the transcriptional level [35]. Similar to the patterns for lignin biosynthetic genes, we found a number of genes were up-regulated in *pal1 pal2*, *ref3-2*, *ref3-3*, *cadC cadD*, *med5a/b*, *mefd5a/b ref8*, *med5a/b ref2*, but down-regulated in *cse-2*, *fah1*, and *ref2* (Figure 4.9). *pal1 pal2* showed slight increase in the expression of 3-DEOXY-D-ARABIONO-HEPTULOSONATE 7-PHOSPHATE SYNTHASE 2 (*DAHPS2*), AROGENATE DEHYDRATASE 2 (*ADT2*), and *ADT3*. *cadC cadD*, similar to *pal1 pal2*, had modest enhancement in expression of *DAHPS2*, 3-DEHYDROQUINATE SYNTHASE (*DHQS*), SHIKIMATE KINASE 1 (*SK1*), and four of the six *ADT* genes including the major isoform *ADT4* and *ADT5* [36], but had a log₂FC of -0.2 for *ADT3* and *CHORISMATE MUTASE 2 (CM2)*. *ref3-2* exhibited more substantial changes than *ref3-3* for the three *DAHPS* genes and *SK2*, and the major *ADT* genes. *med5a/b*, *med5a/b ref8*, and *med5a/b ref2* had up-regulation in the genes responsible for the synthesis of chorismate and the four *ADT* genes that were also up-regulated in *cadC cadD*, whereas the three mutants all showed decreased transcripts of *ADT1*, an isoform with a lower FPKM than *ADT4* and *ADT5* in wild type. On the other hand, *cse-2*, *fah1*, and *ref2* displayed reduced expression in genes involved in every step except chorismate synthase (*CS*). Despite the slight up-regulation of *DAHPS2* and *ADT2* in both *fah1* and *ref2*, the repressed transcription of the shikimate biosynthetic genes in the three mutants was consistent with their suppression of the lignin biosynthetic genes. These results suggest that transcription of genes for Phe biosynthesis and subsequent lignification is coordinated in the rapidly lignifying stem tissue.

The global transcriptional changes in lignin biosynthetic genes seen in each mutant might be a result of differential expression of the TFs that simultaneously regulate these genes. We thus examined the transcript abundance of the characterized TFs involved (Figure 4.10) [5-7]. Altered expression of the known TFs was observed, likely contributing to the changed expression of the lignin biosynthetic genes. *cse-2*, *fah1*, and *ref2* showed repression in the activators such as *MYB42*, *MYB58*, *MYB63*, and *MYB85*, but activation of the repressor *MYB4* except in *fah1*. In contrast, in *ref3-2*, *ref3-3*, *C4H:F5H*, and *ref8 fah1 SmF5H*, increased transcript abundance was observed for numerous transcription activators, but decreased mRNA levels for *MYB4*. The opposite alteration patterns of these TFs possibly led to disparate changes in the biosynthetic genes in *cse-2*, *fah1*,

ref2 versus *ref3-2*, *ref3-3*, *cadC cadD*, *C4H:F5H*, and *ref8 fah1 SmF5H*. *MED5a* and *MED5b* as co-regulators repress the transcription of genes in lignin biosynthesis. Subtle changes in their expression were seen in these biosynthetic mutants (Figure 4.10). Interestingly, in *ref3-2*, *cadC cadD*, and *C4H:F5H*, the only three biosynthetic mutants in which the paralogous genes were statistically altered in expression, *MED5b* was down-regulated but *MED5a* was up-regulated.

KFB1, *KFB20*, *KFB30* and *KFB50* mediate the degradation of PAL to post-translationally control the gateway for phenylpropanoid metabolism [10, 11]. We found that expression of the *KFB* genes also differed from wild type in the mutants (Figure 4.10). In *pal1 pal2*, *KFB39* was substantially repressed, potentially alleviating the breakdown of PAL enzymes. *KFB39* as well as *KFB50* was also down-regulated in the two *ref3* mutants, *4cl1*, and *C4H:F5H*. More substantial decreases in *KFB39* and *KFB50* were seen when *MED5a* and *MED5b* were knocked-out (Figure 4.10) and the mutants accumulated higher levels phenylpropanoids (Figure 4.7), consistent with published data from *med5a/b* rosette [33]. Besides the down-regulation of *KFB39* and *KFB50*, *KFB01* was up-regulated in *med5a/b ref8* and *med5a/b ref2*, and *KFB20* in *med5a/b ref2* compared to wild type, respectively. *cse-2*, with reduced expression of the *PAL* genes, had over two-fold increase in the transcripts of *KFB39* (Figure 4.10). On the other hand, *fah1*, which also had lower expression of the *PAL* genes, showed reduced transcription of all four *KFB* genes. These results demonstrated that the *KFB* genes respond differently in these mutants.

4.2.4 Analysis of differentially expressed genes (DEGs) in phenylpropanoid mutants identified multiple mis-regulated processes

The analysis described above revealed that perturbations in the phenylpropanoid mutants have system-wide effects on phenylpropanoid metabolism and the upstream shikimate pathway. To investigate the global impact of changes in phenylpropanoid metabolism on the transcriptome in our mutants, we compared the expression of all 20974 expressed genes in our dataset and identified genes with at least a 2-fold increase or decrease (false discovery rate (FDR) < 0.01) in each mutant (Figure 4.11). In aggregate, 5581 DEGs were identified from the 13 mutants. The mutants that scattered closer to wild type on PCA plot (Figure 4.3) had fewer than 1000 DEGs, including *pal1 pal2*, *4cl1*, *cse-2*, *fah1*, and *ref2*, whereas, over 1000 DEGs were observed in the other mutants. Overexpression of *F5H* caused about 2500 DEGs in the stems, more than all of the loss-of-function biosynthetic mutants included. In the three plants with *MED5a/b* knocked out,

about 2000 genes were mis-regulated, more than the numbers reported before from the rosette of *med5a/b* and *med5a/b ref8* [9].

To quantitatively examine these DEGs among the mutants, we performed Pearson's correlation analysis using the log₂FC of the 5581 mis-regulated genes (Figure 4.12). *pal1 pal2*, *4cl1*, *cse-2*, *fah1*, and *ref2* not only showed smaller numbers of DEGs, but also exhibited positively correlated DEG profiles, which suggested that common DEGs in pair-wise comparison of the five mutants overall had changes of the same direction. On the other hand, the mutants with more DEGs also showed positively correlated DEG profiles among themselves, suggesting that shared DEGs between every two mutants were generally mis-regulated in the same direction. Intriguingly, the global changes in gene expression in *fah1*, and *ref2* to a lesser extent, negatively correlated with that in the group harboring *C4H:F5H*. The patterns were revealed with details of DEGs when hierarchical clustering analysis was conducted to group both mutants and the 5581 DEGs (Figure 4.13). The five mutants with fewer DEGs separated from the other eight with more DEGs in the clustering. Based on the changes in their gene expression the former further formed subgroup A and B, and the latter formed subgroup C, D, and E. *fah1* and *ref2* grouped tightly, with decreased transcripts for gene Cluster 2-8 and increased transcripts for gene Cluster 13-15. Genes in these clusters exhibited opposite changes in the eight mutants with higher numbers of DEGs, resulting in a negative correlation in gene expression profiles with *fah1-2* and *ref2*. As seen before in Figure 4.12, DEG profiles were alike in *ref3-2* and *ref3-3*, with more substantial changes in the stronger allelic mutant *ref3-2* (Figure 4.13). The eight mutants in the subgroup C, D, and E had remarkably similar changes in DEGs. Up-regulated genes included those involved in various defense responses and salicylic acid (SA), jasmonic acid (JA), and abscisic acid (ABA) signaling, and repressed genes included those responsible for photosynthesis, lipid catabolism, auxin signaling (Table 4.3, Figure 4.13). It appeared that growth related genes were suppressed in these perturbed plants while genes associated with responses to biotic and abiotic stresses were induced, although the plants were grown in a growth chamber under normal, controlled conditions and did not encounter stresses.

4.2.5 Comparison of metabolites and DEGs in subgroups of mutants

Five subgroups with comparable DEG profiles emerged from the hierarchical clustering analysis (Figure 4.13). To further analyze their commonly mis-regulated genes and correlate

transcript alterations with metabolite accumulation, we compared the measured phenylpropanoids and DEGs with at least 2-fold changes in wild type and the mutants within a series of subgroups.

pal1 pal2, fah1, and ref2

pal1 pal2, fah1, and ref2 constituted the first subgroup. Examination of their DEG profiles again revealed an overlap of genes with alteration of the same direction in the three mutants, and a more substantial overlap of down-regulated genes in *fah1* and *ref2* (Figure 4.14A, Figure 4.14B). Functional analysis of the common DEGs showed carbon fixation enriched in the 14 up-regulated genes but no significant enrichment (FDR < 0.01) in the 44 down-regulated genes (Table 4.4). We analyzed the 338 repressed genes shared by *fah1* and *ref2* and found significant enrichment for genes involved in transcription regulation and response to chitin and cold (Table 4.4). Among these genes was *KFB* targeting chalcone synthase [12], the enzyme that catalyzes the first committed step for flavonoid production, indicating altered post-translational regulation for flavonoid metabolism in *fah1* and *ref2*.

All three mutants in this subgroup showed lower levels of hydroxycinnamyl aldehyde intermediates and sinapyl alcohol (Figure 4.14C). The reduction in sinapaldehyde and sinapyl alcohol is consistent with lower levels of S lignin in *fah1* and *ref2* reported before [23, 26]. Transcript abundance of *F5H*, *COMT1* and *CAD5* was significantly decreased in *ref2* compared to wild type (Figure 4.8), which may cause reduced enzyme activities and subsequent lower levels of products. This model is consistent with the observations that *fah1* and *ref2* accumulated ferulate. The fact that the immediate substrate of F5H, coniferaldehyde, does not accumulate suggests that re-distribution of flux occurs by oxidation of the hydroxycinnamaldehyde to the hydroxycinnamate possibly by the hydroxycinnamaldehyde dehydrogenase encoded by *REF1* [37].

ref3-2 and ref3-3

ref3-2 is a stronger allele of *C4H* than *ref3-3* [22]. *ref3-2* mutants showed stunted growth whereas *ref3-3* did not (Figure 4.1). As mentioned before in Figure 4.11 and Figure 4.13, *ref3-2* had more up and down regulated DEGs and DEGs with larger changes compared to *ref3-3* (Figure 4.15A, Figure 4.15B). Functional analysis of the overlapping genes with increased transcription revealed enrichment for responses to protein phosphorylation, SA, defense, and flavonoid glucuronidation. Among those genes were *DOWNY MILDEW RESISTANT 6 (DMR6)* and *DMR6-LIKE OXYGENASE 1*, encoding SA 5-hydroxylase and SA 3-hydroxylase, respectively, which have been shown to be induced by SA treatment and maintain SA homeostasis in plants [38, 39]. The

387 common down-regulated genes showed enrichment in cell wall organization, including a number of xyloglucan:xyloglucosyl transferase genes and expansin genes (Table 4.4).

Both *ref3* mutants showed decreased levels of *p*-coumaraldehyde, coniferaldehyde (Figure 4.15C). In addition, the content of caffeoyl CoA and sinapaldehyde were lower in *ref3-2*, but not in *ref3-3*. The reduction in these pathway intermediates was consistent with the lower lignin phenotype of *ref3* mutants reported before [22].

4cl1 and *cse-2*

Both *4cl1* and *cse-2* mutants were morphologically like wild type when we sampled the basal stems (Figure 4.1), although *cse-2* has been shown a dwarf phenotype at maturity [21]. Both *4cl1* and *cse-2* mutants had more down-regulated DEGs than up-regulated DEGs (Figure 4.16A, Figure 4.16B). The majority of common down-regulated genes in *4cl1* and *cse-2* were involved in stress responses, including a number of *WRKY*, *ERF*, *CBF* transcription factors (Table 4.4).

4cl1 was marked by a much higher level of *p*-coumarate (Figure 4.16C), consistent with over 80% reduction in 4CL activity when this major isoform gene is disrupted [14]. *cse-2* had a distinct hyper-accumulation of the substrate for the enzyme defective in the mutant, caffeoyl shikimate (Figure 4.16C). In *4cl1* and *cse-2*, coniferaldehyde and sinapaldehyde levels were lower, however the corresponding ferulate and sinapate levels were much higher. This result suggested oxidation of the hydroxycinnamaldehydes to the less active hydroxycinnamates, similar to the re-distributed flux seen in the *fah1* and *ref-2* mutants. Surprisingly, *REF1*, encoding the aldehyde dehydrogenase that catalyzes the conversion [37], had no change in expression in either mutant compared to wild type (Figure 4.8). When one-way ANOVA was done with all 14 genotypes for each metabolite, concentrations of *p*-coumaraldehyde and *p*-coumaryl alcohol were higher in *cse-2* (Table 4.2), consistent with an increase in H lignin synthesis reported previously [21].

C4H:F5H and *cadC cadD*

The *C4H* promoter was very effective in driving the overexpression of *F5H* such that its transcript was about 60 times that of wild type and *F5H* became the most highly expressed gene in the stem (Figure 4.8). Interestingly, *cadC cadD* and *C4H:F5H* both displayed large numbers of DEGs compared to wild type, and shared a substantial fraction of mis-regulated DEGs (Figure 4.17A, Figure 4.17B). Further functional analysis revealed that the common up-regulated genes in *cadC cadD* and *C4H:F5H* showed enrichment for responses to detoxification, and the common

down-regulated genes were related to auxin response and its polar transport, cell wall biogenesis and organization (Table 4.4).

CAD and F5H compete for the common substrate coniferaldehyde in plant cells. When the major isoforms of *CAD* were knocked out in *cadC cadD*, the only changes observed in metabolite content were enhancement in coniferaldehyde and reduction in coniferyl alcohol (Figure 4.17C). The changes are consistent with the deficiency of CAD activity. In *C4H:F5H*, coniferyl alcohol level was lower than wild type, but no significant difference was seen for other phenylpropanoids when the comparison was done within the subgroup (Figure 4.17C).

med5a/b, med5a/b ref8, med5a/b ref2

Disruption of *MED5a/b* results increased transcripts of lignin biosynthetic genes and enhanced levels of phenylpropanoids in rosettes [9, 33]. We observed the same trend for transcripts and intermediates for lignin biosynthesis in the stem tissue from *med5a/b*, *med5a/b ref8*, and *med5a/b ref2* (Figure 4.18). All three mutants showed large numbers of DEGs compared to wild type, and shared a majority of them, being dictated by the lack of *MED5* (Figure 4.11 and Figure 4.18A). An enrichment in SA response and various defense responses towards external stresses was identified from functional analysis of the common up-regulated genes. Among the 440 genes were *DMR6* and *DMR6-LIKE OXYGENASE 1*, that were also up-regulated in the *ref3* mutants. Analysis of the shared down-regulated genes showed enrichment in auxin regulated growth process, cell wall and cell membrane biogenesis and degradation. A significant number of these genes encode secreted proteins, including glycoproteins and expansins (Table 4.4).

In contrast to most of the phenylpropanoids with increased accumulation, Phe content was reduced by about 50% (Figure 4.18C) and shikimate pathway was transcriptionally activated (Figure 4.9) in these three mutants. The decrease in Phe content is consistent with the suppressed expression of *KFB39* and *KFB50*. The *med5a/b* double mutant had increased levels of *p*-coumaryl alcohol, coniferyl alcohol and sinapyl alcohol, however the three *LACCASE* genes and *PRX64* involved in lignification showed decreased transcript levels (Figure 4.8), which could be a reason that *med5a/b* does not deposit more lignin than wild type [9]. More substantial reductions in the *LACCASE* genes and *PRX64* was seen in *med5a/b ref2*, consistent with its hyper-accumulation of *p*-coumaryl alcohol and coniferyl alcohol. Like *cse-2*, *med5a/b ref8* exhibited higher accumulation of *p*-coumaryl derivatives because *C3'H* was blocked, consistent with a higher H lignin deposition reported before (Figure 4.18C) [9].

fah1, *C4H:F5H*, and *ref8 fah1 SmF5H*

In addition to the subgroups derived from the hierarchical analysis, we also compared *fah1*, *C4H:F5H*, and *ref8 fah1 SmF5H* together with wild type because they all have perturbations at the F5H step (Figure 4.19). As expected, *F5H* transcript was absent in *fah1* but highly expressed in *C4H:F5H* (Figure 4.8). *ref8 fah1 SmF5H* is deficient in *C3'H* and *F5H*, but expresses an exogenous *F5H* whose encoded enzyme catalyzes both 3- and 5- hydroxylation. The mutant accumulates mainly H and S lignin [25]. As seen in Figure 4.12, a majority of shared DEGs between *fah1* and *C4H:F5H* showed changes of opposite directions (Figure 4.19A). The same trend was observed between *fah1* and *ref8 fah1 SmF5H*. In contrast, common DEGs between *C4H:F5H* and *ref8 fah1 SmF5H* overall were mis-regulated to the same direction (Figure 4.19A). We thus examined the DEGs with decreased expression in *fah1* and increased in *C4H:F5H* and *ref8 fah1 SmF5H*, and the DEGs with the reverse pattern for functional analysis. We found enrichment for genes involved in extracellular processes and secretion (Table 4.4).

Consistent with the mutation in *C3'H*, higher levels of *p*-coumarate, *p*-coumaroyl shikimate, and *p*-coumaraldehyde were accumulated in *ref8 fah1 SmF5H* (Figure 4.18C). In addition, *ref8 fah1 SmF5H* had substantial increase of sinapate in spite of a lower transcript level of *REF1* that encode the aldehyde dehydrogenase (Figure 4.8).

4.3 Discussions

Plants tightly control lignin biosynthesis at multiple levels to allocate carbon towards this significant carbon sink. Despite this quantitative control, there is a great deal of qualitative flexibility in monomer composition, given that available phenylpropanoid intermediates such as coniferaldehyde, sinapaldehyde, *p*-coumaryl alcohol, coniferyl alcohol, and sinapyl alcohol can function as primary lignin subunits in genetically modified plants [9, 17, 20, 40]. In addition to the bioengineered lignin, modifications of lignin biosynthesis lead to global alterations in transcription and metabolite profiles [13, 19]. In this study, we conducted genome-wide transcriptome analysis and phenylpropanoid profiling in *Arabidopsis* plants of 14 different genotypes to understand how plants respond to genetic perturbations of lignification. Our molecular phenotyping revealed substantial changes at the transcriptional and metabolic levels in mutants with perturbed lignin biosynthesis, even when there was no obvious growth difference. The transcript and

phenylpropanoid measurements also provided *in planta* data to simulate metabolic flux and how the control in these mutants is perturbed with altered metabolism.

Consistent with many previous reports [9, 13, 19], we found that genetic manipulation at each step of lignin biosynthesis affected the expression of other genes in the pathway. Interestingly, most of the genes in the shikimate pathway were mis-regulated in the same direction as lignin biosynthetic genes in a number of our mutants, including *ref3-2*, *ref3-3*, *cse-2*, *fah1*, *med5a/b*, *med5a/b ref8*, and *med5a/b ref2* (Figure 4.8, Figure 4.9). These observations suggest that the upstream supply of Phe and lignification are tightly coordinated in plants at transcriptional level. Similarly, up-regulation of genes in the shikimate and flavonoid pathways has been reported in transgenic tomato fruits that express the Arabidopsis TF *MYB12* [41]. In another instance, consistent suppression of the shikimate and phenylpropanoid pathway genes occurs in Arabidopsis *nst1 nst3* mutant that is deficient in the master TFs regulating secondary cell wall formation [42]. These examples suggest that coregulation of the genes for Phe biosynthesis and lignification involves TFs [43]. Indeed, in our RNA-seq data the expression of *MYB42*, an activator TF [44], was decreased in *fah1* and the expression of *MYB4*, a repressor TF [44], increased in *cse-2*, consistent with the repression of the biosynthetic genes in both mutants (Figure 4.10). In addition, the increased transcripts of *MYB42* and decreased of *MYB4* were in consistency with the induction of the pathway genes in *ref3* mutants and the *med5a/b* containing mutants. In these five mutants, *NST1* nevertheless showed decreased expression (Figure 4.10), suggesting the activation of the pathway genes is independent of the mis-regulation of *NST1*.

In biosynthetic mutants with altered metabolite accumulation, genes in the same pathway that are not mutated are often up-regulated [9, 13]. Likewise, *pal1 pal2*, *ref3-2*, *ref3-3*, *4cl1*, *cadC* *cadD* showed activation of the phenylpropanoid genes. In contrast, *cse-2* and *fah1* exhibited reduced expression of the shikimate and phenylpropanoid biosynthetic genes (Figure 4.8, Figure 4.9). In *cse-2*, the decreased gene expression was consistent with reduced content of coniferaldehyde and sinapaldehyde (Figure 4.15), and lower level of total lignin reported [21]. On the contrary, although *fah1* showed reduced transcript abundance for genes in nearly every step of the shikimate and lignin biosynthetic pathways and decreased accumulation of a number of phenylpropanoids including the three aldehyde intermediates and sinapyl alcohol (Figure 4.13), its total lignin content stayed the same as wild type [19]. These results suggest in *fah1* expression of shikimate and phenylpropanoid genes does not determine the total carbon flux channeled towards

lignin. Reduced accumulation of hydroxycinnamates and anthocyanins in *fah1* has been reported previously and has been suggested as a result of accumulation of some phenylpropanoids between C3'H and F5H that represses the pathway via a Mediator-dependent feedback mechanism [45]. These observations bring up the question: what metabolites cause the opposite transcriptional changes of the shikimate and phenylpropanoid pathway genes in *cse-2* and *fah1* versus in *pal1 pal2*, *ref3-2*, *ref3-3*, *4cl1*, and *cadC cadD*? In our targeted LC-MS/MS analysis, no common phenylpropanoid was altered uniquely to both *cse-2* and *fah1* (Table 4.2), but it remains possible that some derivatives of phenylpropanoids not measured here are involved as regulators. Alternatively, different metabolites may be responsible for the repression in *cse-2* and *fah1*, respectively, which is in line with the distinct mis-regulation of *MYB42* in *cse-2* and *MYB4* in *fah1*.

In addition to the metabolites leading to suppression of phenylpropanoid metabolism in *fah1* and *cse-2*, multiple other Phe-derived compounds with altered accumulation are likely involved in affecting gene expression. The *F5H* over-expresser, *C4H:F5H*, shared a remarkably similar transcription profile with *cadC cadD* (Figure 4.13), which is possibly triggered by compounds that are commonly altered in the two genotypes. Both *C4H:F5H* and *cadC cadD* exhibited reduction in coniferyl alcohol and hyper-accumulation of sinapaldehyde (Figure 4.17, Table 4.2), making these two compounds possible mediators of changed gene expression. Dehydrodiconiferyl alcohol glucoside (DAG), a product of coniferyl alcohol, is suggested to promote cell division and expansion [46-48]. If DAG indeed controls cell growth as a regulator, it must almost certainly act by altered gene expression, like the other plant hormones. The lower level of coniferyl alcohol in *C4H:F5H* and *cadC cadD* potentially results in a reduction in DAG, thus leads to similar DEG patterns. The down-regulation of expansin and cell wall organization genes in the two perturbed plants also supports this possibility (Table 4.4). Similar to *C4H:F5H* and *cadC cadD*, *ref3* mutants and the *med5a/b* containing mutants also exhibited high similarity in their DEG profiles (Figure 4.13), marked by up-regulation of genes involved in SA responses. The activation of SA responsive genes is consistent with higher SA content in *med5a/b ref8* rosette [9]. Hyperaccumulation of SA has also been reported in *Artemisia annua* with RNAi suppressed *C4H* [49] and in tobacco cell culture treated with C4H inhibitors [50]. Blockage in C4H causes accumulation of cinnamate and thus higher synthesis of downstream product SA, which is likely the same for *ref3-2* and *ref3-3*. Interestingly, in these five mutants SA 5-hydroxylase and SA 3-hydroxylase encoding genes, known to be induced by high SA [38, 39], showed increased

expression, suggesting catabolism of SA to maintain SA homeostasis. The evidence together proposes a model in which the *ref3* mutants and *med5a/b* containing mutants hyperaccumulate SA which triggers induction of SA responsive genes and leads to similar DEG profiles.

For the lignin biosynthetic steps involving multiple genes isoforms, namely *PAL*, *4CL*, *CCR*, *CAD*, and *LAC*, some isoforms showed constantly high expression, while others had relatively low expression in wild type but responded sensitively in these mutants (Figure 4.8). More specifically, we found that *PAL1*, *4CL1*, *CCR1*, and *LAC4* had higher transcript abundance than their paralogs in wild type and showed subtle changes in expression in mutants in which they are not mutated. In contrast, *PAL2*, *4CL2*, *4CL4*, *CCR2*, *LAC11*, and *LAC17* showed lower expression in wild type than the former four isoforms, respectively, and exhibited more substantially altered expression in a number of the mutants (Figure 4.8, Table 4.1). Similar observation has been reported for *CCR* genes in response to external stimuli. *CCR1* shows much higher expression than *CCR2* in Arabidopsis stem under normal condition, but is not induced by pathogen inoculation, whereas *CCR2* is expressed at only low levels during development but is strongly activated under stresses [51]. Both the genetic manipulations and external stresses trigger the induction of these gene isoforms, which suggests that these genetic perturbations and environmental stimuli stimulate gene regulation via shared molecular machinery, such as the TFs involved.

Four KFB proteins redundantly regulate the turnover of PAL enzymes [10, 11]. The expression of *KFB39* and *KFB50* were substantially decreased in a few of the perturbed plants, including *pal1 pal2*, *ref3-2*, *ref3-3*, *4cl1*, *fah1*, *C4H:F5H*, *med5a/b*, *med5a/b ref8*, and *med5a/b ref2*, suggesting that they were regulated similarly to diminish the degradation of PAL (Figure 4.10). Repression of the *KFB* genes is synchrony with the enhanced expression of the phenylpropanoid pathway genes to increase metabolic flux into the pathway, given that production of end products such as lignin and hydroxycinnamate derivatives is reduced in most of the biosynthetic mutants [9, 13, 14, 20, 22]. Consistent with the repressed phenylpropanoid gene expression, *cse-2* showed increased transcript level of *KFB39* (Figure 4.10), which is distinct from *pal1 pal2*, *ref3-2*, *ref3-3*, and *4cl1*. This difference suggests that the accumulation of metabolites downstream of 4CL but upstream of CSE (e.g. *p*-coumaroyl shikimate, or caffeoyl shikimate) leads to the opposing transcriptional changes of *KFB39*, and possibly the shikimate and phenylpropanoid biosynthetic genes as well. Significant reduction in *KFB39* and *KFB50* expression in *med5a/b*,

med5a/b ref8, and *med5a/b ref2* suggests that MED5a/b functions as a coactivator for their transcription. These results together indicate that post-translational regulation of PAL enzymes by KFBs is coordinated with the existence of the homeostatic mechanisms to maintain phenylpropanoid metabolic flux.

We can assemble the transcriptomic and metabolomic data to gather qualitative information about the metabolic activities for phenylpropanoid network in these perturbed plants. Here we include four examples. The first example is *fahl*, which has been proposed to accumulate some phenylpropanoids between C3'H and F5H that lead to anthocyanin deficiency, independent of changed expression of the biosynthetic genes involved [39]. Actually, we found that *CHS*, as well as *4CL3* with a predominant role in flavonoid biosynthesis [14], showed increased expression in *fahl* compared to wild type. This result led us to hypothesize that although the transcript abundance of *CHS* is higher in *fahl*, degradation of CHS protein is also enhanced by KFB^{CHS} and thus results in lower levels of anthocyanins. The transcript level of *KFB^{CHS}*, however, was decreased in *fahl* (Table 4.4), potentially leading to higher CHS activity. These data support the idea that its deficiency in anthocyanin accumulation does not involve altered gene expression or post-translational modification. The second example is *pal1 pal2*, which deposits lignin of 35% wild-type level [13] and accumulates Phe of about 25-fold wild-type level (Table 4.2). The expanded Phe pool in *pal1 pal2* is constant in the rapidly lignifying young stems (data not shown), which suggests the synthesis rate of Phe in *pal1 pal2* is reduced to about 35% of that in wild type (assuming Phe is primarily used to produce lignin). Similarly, decrease in Phe production occurs in petunia flowers with RNAi suppressed *PAL* [52]. The reduction can be achieved by repression of the shikimate pathway genes or post-transcriptional regulation on the enzymes. The first possibility is ruled out since no significant decrease was observed in the expression of the genes for Phe biosynthesis in *pal1 pal2* (Figure 4.9). Instead, expression of *DAHPS2*, *ADT2*, and *ADT3* were slightly enhanced, consistent to previous reports in *pal1 pal2* of Arabidopsis ecotype C24 [13] and in *RNAi-PAL* suppressed petunia flowers [52]. These results suggest that hyper-accumulation of Phe leads to reduced shikimate pathway flux without transcriptional suppression of the pathway genes, as previously proposed in the petunia flowers study [52]. Third, in contrast to *pal1 pal2*, the three mutants with *MED5a/b* knocked out had increased transcript abundance of genes involved in every step of the shikimate pathway (Figure 4.9), suggesting an enhanced capacity for higher flux. Instead of a higher level of Phe, *med5a/b*, *med5a/b ref8*, and *med5a/b*

ref2 contained approximately half of Phe content in wild type, and increased levels of downstream phenylpropanoids (Figure 4.18). In addition, the *PAL* genes were up-regulated and *KFB39* and *KFB50* were down-regulated in these mutants, potentially leading to higher PAL activity driving flux towards the phenylpropanoid pathway. The increase in PAL activity and induction in Phe supply in our previous report [53] are complementary strategies to enhance the phenylpropanoid flux in plants. Based on these data we can infer fast turnover of Phe for phenylpropanoid production via an integration of transcriptional and post-translational regulation in the absence of MED5a/b. The results in turn indicate that *MED5a/b* mediates the homeostasis of phenylpropanoid metabolism through coherent repressive mechanisms in wild type plants. Last, the datasets are also informative for the qualitative estimation of flux allocation through branches in the pathway. CSE together with 4CL bypasses HCT for the conversion of caffeoyl shikimate to caffeoyl CoA [21]. The substantial increase in caffeoyl shikimate (Table 4.2) and 36% decrease in lignin in *cse-2* [21] suggest that the flux through HCT is relatively low, which might be, in part, contributed by the decreased expression of *HCT* in *cse-2*. To evaluate and test these qualitative inferences of flux shifts in phenylpropanoid metabolism, mathematical simulation based on the data is necessary to calculate the accurate metabolic activities and pinpoint the controlling factors [27].

4.4 Methods

4.4.1 Plant materials and growth

Plant lines in this study are all in *Arabidopsis thaliana* Columbia-0 background. *pall* (SALK_096474C), *pal2* (GABI_692H09-025071), *4cl1* (WiscDsLox473B01), *cse-2* (SALK_023077), *cadC* (SALL_1265_A06), *cadD* (SALL_776_B06), *med5a* (SALK_011621), *med5b* (SALK_037472), are T-DNA insertional mutants obtained from ABRC. *ref3-2*, *ref3-3*, *ref2*, *ref8*, *fah1* are mutants isolated from ethyl methanesulfonate-generated population with reduced accumulation of sinapoylmalate as reported before [22, 24, 54]. *F5H* over-expresser (*C4H:F5H*) is a transgenic plant with *C4H* promoter driven *Arabidopsis F5H* [23]. *SmF5H* contains a *Selaginella moellendorffii F5H* driven by *Arabidopsis C4H* promoter [25]. Higher order mutants were made by crosses of the single mutants. For the mutants with point mutations, restriction enzyme digestion of the PRC products was performed to confirm the genotype. Amplicons from *ref3-2* was not digested by *HinfI* while from wild type was cleaved. Amplicons from *ref3-3* was

not cleaved by BamHI while from wild type was. The PCR products from *fah1-2* mutant was digested by MseI into four fragments while that from wild type was digested into three fragments. The amplicons from *ref8* were digested by EcoRV while the amplicon from wild type was not cut. Genotypes of these mutants were confirmed with PCR. Primer sequences for genotyping are listed in Table 4.5.

Arabidopsis thaliana plants were grown in soil in one growth chamber at 23 °C under 200 $\mu\text{E m}^{-2} \text{s}^{-1}$ with 16-hour light/ 8-hour dark cycle. The basal 0.5-2 cm of the inflorescent stems were harvested and flash frozen in liquid nitrogen. At noon 28 days after planting, 6 plants from 6 randomly positioned pots were gathered for one biological sample for RNA-seq experiment. Four samples collected for each mutant genotype and eight samples for wild type. At noon on day 29, 10 plants from six randomly positioned pots were collected as one replicate for soluble metabolites analysis with three replicates for each genotype. Samples were stored at -80 °C before extraction for total RNA or soluble metabolites.

4.4.2 RNA-seq

Frozen stem samples were ground in ceramic mortars and pestles in liquid nitrogen. About 100 mg ground tissue per sample were used for RNA extraction with the QIAGEN RNeasy Mini kit. An on-column DNase treatment was performed to remove DNA from samples. Total RNA samples were submitted to the Purdue Genomics Core Facility for library construction and subsequent sequencing. PolyA+ RNA libraries were constructed using Illumina's TruSeq stranded mRNA library prep kit for each sample. Double-end 100 bp sequencing were done on a single chip with an Illumina's NovaSeq instrument. The raw transcript reads were mapped using the TAIR10 genome build and the HISAT2 alignment program [55].

4.4.3 Statistical analysis of RNA-seq data

The mapped reads from the RNA-seq experiment was utilized to determine digital gene expression (counts) for every exon using TAIR10 gene model and HTSeq-count program [56]. Counts were summarized by gene ID. The counts for each gene in all 60 samples were processed with DESeq2 for differential gene expression analysis [57, 58]. A gene was considered to express at low levels if both its average raw counts were fewer than 30 in at least one genotype and its averaged normalized counts in 60 samples were fewer than 5. Removal of genes with low

expression resulted in a count table comprising 20974 expressed genes. Using the result function in DESeq2, expression of genes in every mutant genotype was compared with eight wild-type controls to identify DEGs (FDR < 0.01). Log₂FC (mutant to wild type) values of 5581 DEGs with at least 2-fold changes in at least one mutant were used for hierarchical clustering analysis. Ward.D2 clustering method was applied to generate the heatmap using gplots package [59]. Gene ontology analysis was done using the DAVID Bioinformatics Resource v6.8 and 20974 expressed genes as background.

4.4.4 Soluble metabolite analysis on LC/MS-MS

Soluble phenylpropanoids were analyzed on LC-MS/MS using the method developed previously [30]. Stem tissue were ground in liquid nitrogen into fine powder then extracted with (10 µL mg⁻¹ fresh weight) 75% methanol: H₂O, for 2 hours at 4 °C. Samples were centrifuged for 20 min at 4000 rpm at 4 °C. The supernatant was dried under nitrogen gas before being re-dissolved in 60 µL 50% Methanol: H₂O. Samples were run on LC-MS/MS for quantification of the phenylpropanoid metabolites. Separation was done on a Zorbax Eclipse C8 column (150 mm 4.6 mm, 5 µm, Agilent Technologies, Santa Clara, CA) with ammonium acetate (pH 5.6) and acetonitrile/H₂O/HCOOH (9.8/2/0.2) as mobile phase. Detection was done on an AbSciex QTrap 5500 triple quadrupole mass spectrometer equipped with an electrospray ionization probe in the negative ion mode. Standards were run with the same methods to obtain the calibration curve.

4.5 References

1. Bonawitz ND, Chapple C: The genetics of lignin biosynthesis: connecting genotype to phenotype. *Annu Rev Genet* 2010, 44:337-363.
2. Barros J, Serrani-Yarce JC, Chen F, Baxter D, Venables BJ, Dixon RA: Role of bifunctional ammonia-lyase in grass cell wall biosynthesis. *Nat Plants* 2016, 2(6):16050.
3. Berthet S, Demont-Caulet N, Pollet B, Bidzinski P, Cezard L, Le Bris P, Borrega N, Herve J, Blondet E, Balzergue S, et al.: Disruption of *LACCASE4* and *17* results in tissue-specific alterations to lignification of *Arabidopsis thaliana* stems. *Plant Cell* 2011, 23:1124-1137.
4. Zhao Q, Nakashima J, Chen F, Yin Y, Fu C, Yun J, Shao H, Wang X, Wang ZY, Dixon RA: LACCASE is necessary and nonredundant with PEROXIDASE for lignin polymerization during vascular development in *Arabidopsis*. *Plant Cell* 2013, 25:3976-3987.

5. Zhong R, Lee C, Zhou J, McCarthy RL, Ye ZH. A battery of transcription factors involved in the regulation of secondary cell wall biosynthesis in *Arabidopsis*. *Plant Cell* 2008, 20:2763-2782.
6. Taylor-Teeple M, Lin L, de Lucas M, Turco G, Toal TW, Gaudinier A, Young NF, Trabucco GM, Veling MT, Lamothe R, et al.: An Arabidopsis gene regulatory network for secondary cell wall synthesis. *Nature* 2015, 517:571-U307.
7. Zhong RQ, McCarthy RL, Lee C, Ye ZH: Dissection of the transcriptional program regulating secondary wall biosynthesis during wood formation in poplar. *Plant Physiol* 2011, 157:1452-1468.
8. Bonawitz ND, Soltan WL, Blatchley MR, Powers BL, Hurlock AK, Seals LA, Weng JK, Stout J, Chapple C: REF4 and RFR1, subunits of the transcriptional coregulatory complex mediator, are required for phenylpropanoid homeostasis in Arabidopsis. *J Biol Chem* 2012, 287:5434-5445.
9. Bonawitz ND, Kim JI, Tobimatsu Y, Ciesielski PN, Anderson NA, Ximenes E, Maeda J, Ralph J, Donohoe BS, Ladisch M, Chapple C: Disruption of Mediator rescues the stunted growth of a lignin-deficient Arabidopsis mutant. *Nature* 2014, 509:376-380.
10. Zhang X, Gou M, Liu CJ: Arabidopsis Kelch repeat F-box proteins regulate phenylpropanoid biosynthesis via controlling the turnover of phenylalanine ammonia-lyase. *Plant Cell* 2013, 25:4994-5010.
11. Zhang X, Gou MY, Guo CR, Yang HJ, Liu CJ: Down-regulation of Kelch domain-containing F-box protein in Arabidopsis enhances the production of (poly)phenols and tolerance to ultraviolet radiation. *Plant Physiol* 2015, 167:337-U548.
12. Zhang X, Abraham C, Colquhoun TA, Liu CJ: A proteolytic regulator controlling chalcone synthase stability and flavonoid biosynthesis in Arabidopsis. *Plant Cell* 2017, 29:1157-1174.
13. Rohde A, Morreel K, Ralph J, Goeminne G, Hostyn V, De Rycke R, Kushnir S, Van Doorselaere J, Joseleau JP, Vuylsteke M *et al*: Molecular phenotyping of the *pall* and *pal2* mutants of *Arabidopsis thaliana* reveals far-reaching consequences on phenylpropanoid, amino acid, and carbohydrate metabolism. *Plant Cell* 2004, 16(10):2749-2771.

14. Li Y, Kim JI, Pysh L, Chapple C: Four isoforms of arabidopsis 4-coumarate:CoA ligase have overlapping yet distinct roles in phenylpropanoid metabolism. *Plant Physiol* 2015 169(4):2409-2421.
15. Mavandad M, Edwards R, Liang X, Lamb CJ, Dixon RA: Effects of trans-cinnamic acid on expression of the bean phenylalanine ammonia-lyase gene family. *Plant Physiol* 1990, 94:671-680.
16. Vanholme B, Houari IE, Wout Boerjan: Bioactivity: phenylpropanoids' best kept secret. *Curr Opin in Biotechnol* 2019, 56:156-162.
17. Wang P, Dudareva N, Morgan JA, Chapple C: Genetic manipulation of lignocellulosic biomass for bioenergy. *Curr Opin Chem Biol* 2015, 29:32-39.
18. Chen F, Dixon RA: Lignin modification improves fermentable sugar yields for biofuel production. *Nat Biotechnol* 2007, 25(7):759-761.
19. Vanholme R, Storme V, Vanholme B, Sundin L, Christensen JH, Goeminne G, Halpin C, Rohde A, Morreel K, Boerjan W: A systems biology view of responses to lignin biosynthesis perturbations in Arabidopsis. *Plant Cell* 2012, 24(9):3506-3529.
20. Anderson NA, Tobimatsu Y, Ciesielski PN, Ximenes E, Ralph J, Donohoe BS, Ladisch M, Chapple C. Manipulation of guaiacyl and syringyl monomer biosynthesis in an arabidopsis cinnamyl alcohol dehydrogenase mutant results in atypical lignin biosynthesis and modified cell wall structure. *Plant Cell* 2015, 27(8):2195-209.
21. Vanholme R, Cesarino I, Rataj K, Xiao YG, Sundin L, Goeminne G, Kim H, Cross J, Morreel K, Araujo P *et al*: Caffeoyl shikimate esterase (CSE) is an enzyme in the lignin biosynthetic pathway in Arabidopsis. *Science* 2013, 341(6150):1103-1106.
22. Schilmiller AL, Stout J, Weng JK, Humphreys J, Ruegger MO, Chapple C: Mutations in the cinnamate 4-hydroxylase gene impact metabolism, growth and development in Arabidopsis. *Plant J* 2009, 60:771-782.
23. Meyer K, Shirley AM, Cusumano JC, Bell-Lelong DA, Chapple C: Lignin monomer composition is determined by the expression of a cytochrome P450-dependent monooxygenase in Arabidopsis. *Proc Natl Acad Sci U S A*. 1998, 95(12):6619-23.
24. Chapple C, Vogt T, Ellis BE, Somerville CR: An Arabidopsis mutant defective in the general phenylpropanoid pathway. *Plant Cell* 1992, 4(11):1413-1424.

25. Weng JK, Akiyama T, Bonawitz ND, Li X, Ralph J, Chapple C: Convergent evolution of syringyl lignin biosynthesis via distinct pathways in the lycophyte *Selaginella* and flowering plants. *Plant Cell* 2010, 22(4):1033-45.
26. Hemm MR, Ruegger MO, Chapple C: The *Arabidopsis ref2* mutant is defective in the gene encoding CYP83A1 and shows both phenylpropanoid and glucosinolate phenotypes. *Plant Cell* 2003, 15(1):179-194.
27. Guo L, Wang P, Jaini R, Dudareva N, Chapple C, Morgan JA: Dynamic modeling of subcellular phenylpropanoid metabolism in *Arabidopsis* lignifying cells. *Metab Eng* 2018, 49:36-46.
28. Kim JI, Murphy AS, Baek D, Lee SW, Yun DJ, Bressan RA, Narasimhan ML: *YUCCA6* over-expression demonstrates auxin function in delaying leaf senescence in *Arabidopsis thaliana*. *J Exp Bot.* 2011, 62(11):3981-92.
29. von Mering C, Jensen LJ, Snel B, Hooper SD, Krupp M, Foglierini M, Jouffre N, Huynen MA, Bork P: STRING: known and predicted protein-protein associations, integrated and transferred across organisms. *Nucleic Acids Res* 2005, 33(Database issue):D433-7.
30. Jaini R, Wang P, Dudareva N, Chapple C, Morgan JA: Targeted metabolomics of the phenylpropanoid pathway in *Arabidopsis thaliana* using reversed phase liquid chromatography coupled with tandem mass spectrometry. *Phytochem Anal* 2017, 28(4):267-276.
31. Chiba Y, Green PJ: mRNA degradation machinery in plants. *J Plant Bio.* 2009, 52:114.
32. Franke R, Humphreys JM, Hemm MR, Denault JW, Ruegger MO, Cusumano JC, Chapple C: The *Arabidopsis REF8* gene encodes the 3-hydroxylase of phenylpropanoid metabolism. *Plant J* 2002, 30:33-45.
33. Dolan WL, Dilkes BP, Stout JM, Bonawitz ND, Chapple C: Mediator complex subunits MED2, MED5, MED16, and MED23 genetically interact in the regulation of phenylpropanoid biosynthesis. *Plant Cell* 2017, 29(12):3269-3285.
34. Herrero J, Fernández-Pérez F, Yebra T, Novo-Uzal E, Pomar F, Pedreño MÁ, Cuello J, Guéra A, Esteban-Carrasco A, Zapata JM: Bioinformatic and functional characterization of the basic peroxidase 72 from *Arabidopsis thaliana* involved in lignin biosynthesis. *Planta* 2013, 237(6):237-1599.

35. Tzin V, Galili G: The biosynthetic pathways for shikimate and aromatic amino acids in *Arabidopsis thaliana*. *Arabidopsis Book* 2010, 8:e0132.
36. Corea ORA, Ki C, Cardenas CL, Kim SJ, Brewer SE, Patten AM, Davin LB, Lewis NG: Arogenate dehydratase isoenzymes profoundly and differentially modulate carbon flux into lignins. *J Biol Chem* 2012, 287(14):11446–59.
37. Nair RB, Bastress KL, Ruegger MO, Denault JW, Chapple C: The *Arabidopsis thaliana* *REDUCED EPIDERMAL FLUORESCENCE1* gene encodes an aldehyde dehydrogenase involved in ferulic acid and sinapic acid biosynthesis. *Plant Cell* 2004, 16(2):544-554.
38. Zhang Y, Zhao L, Zhao J, Li Y, Wang J, Guo R, Gan S, Liu CJ, Zhang K: *S5H/DMR6* encodes a salicylic acid 5-hydroxylase that fine-tunes salicylic acid homeostasis. *Plant Physiol* 2017, 175(3):1082-1093.
39. Zhang K, Halitschke R, Yin C, Liu CJ, Gan SS: Salicylic acid 3-hydroxylase regulates *Arabidopsis* leaf longevity by mediating salicylic acid catabolism. *Proc Natl Acad Sci U S A* 2013, 110(36):14807-12.
40. Vanholme R, Morreel 3, Darrah C, Oyarce P, Grabber JH, Ralph J, Boerjan W: Metabolic engineering of novel lignin in biomass crops. *New Phytol* 2012, 196:978-1000.
41. Zhang Y, Butelli E, Alseekh S, Tohge T, Rallapalli G, Luo J, Kwar PG, Hill L, Santino A, Fernie AR, Martin C: Multi-level engineering facilitates the production of phenylpropanoid compounds in tomato. *Nat Commun* 2015, 6:8635.
42. Mitsuda N, Iwase A, Yamamoto H, Yoshida M, Seki M, Shinozaki K, Ohme-Takagi M: NAC transcription factors, NST1 and NST3, are key regulators of the formation of secondary walls in woody tissues of *Arabidopsis*. *Plant Cell* 2007, 19(1):270-80.
43. Maeda H, Dudareva N: The shikimate pathway and aromatic amino acid biosynthesis in plants. *Annu Rev Plant Biol* 2012, 63:73-105.
44. Zhong R, Ye ZH: Transcriptional regulation of lignin biosynthesis. *Plant Signal Behav* 2009, 4(11):1028-34.
45. Anderson NA, Bonawitz ND, Nyffeler K, Chapple C: Loss of FERULATE 5-HYDROXYLASE leads to Mediator-dependent inhibition of soluble phenylpropanoid biosynthesis in *Arabidopsis*. *Plant Physiol* 2015, 169(3):1557-67.

46. Binns AN, Chen RH, Wood HN, Lynn DG: Cell division promoting activity of naturally occurring dehydrodiconiferyl glucosides: do cell wall components control cell division? *Proc Natl Acad Sci U S A*. 1987, 84(4):980-4.
47. Orr JD, Lynn DG. Biosynthesis of dehydrodiconiferyl alcohol glucosides: implications for the control of tobacco cell growth. *Plant Physiol* 1992, 98(1):343-52.
48. Tamagnone L, Merida A, Stacey N, Plaskitt K, Parr A, Chang CF, Lynn D, Dow JM, Roberts K, Martin C: Inhibition of phenolic acid metabolism results in precocious cell death and altered cell morphology in leaves of transgenic tobacco plants . *Plant Cell* 1998, 10(11):1801-16.
49. Kumar R, Vashisth D, Misra A, Akhtar MQ, Jalil SU, Shanker K, Gupta MM, Rout PK, Gupta AK, Shasany AK: RNAi down-regulation of cinnamate-4-hydroxylase increases artemisinin biosynthesis in *Artemisia annua*. *Sci Rep* 2016, 6:26458.
50. Schoch GA, Nikov GN, Alworth WL, Werck-Reichhart D: Chemical inactivation of the cinnamate 4-hydroxylase allows for the accumulation of salicylic acid in elicited cells. *Plant Physiol* 2002, 130(2):1022-31.
51. Lauvergeat V, Lacomme C, Lacombe E, Lasserre E, Roby D, Grima-Pettenati J: Two cinnamoyl-CoA reductase (CCR) genes from *Arabidopsis thaliana* are differentially expressed during development and in response to infection with pathogenic bacteria, *Phytochemistry* 2011, 57(7):1187-1195.
52. Lynch JH, Orlova I, Zhao C, Guo L, Jaini R, Maeda H, Akhtar T, Cruz-Lebron J, Rhodes D, Morgan J, *et al*: Multifaceted plant responses to circumvent Phe hyperaccumulation by downregulation of flux through the shikimate pathway and by vacuolar Phe sequestration. *Plant J* 2017, 92(5):939-950.
53. Wang P, Guo L, Jaini R, Klempien A, McCoy RM, Morgan JA, Dudareva N, Chapple C: A ¹³C isotope labeling method for the measurement of lignin metabolic flux in *Arabidopsis* stems. *Plant Methods* 2018, 23:14:51.
54. Ruegger M, Chapple C: Mutations that reduce sinapoylmalate accumulation in *Arabidopsis thaliana* define loci with diverse roles in phenylpropanoid metabolism. *Genetics* 2001, 159(4):1741-9.
55. Kim D, Langmead B, Salzberg SL: HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* 2015, 12(4):357-60.

56. Anders S, Pyl PT, Huber W: HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* 2015, 31(2):166-9.
57. R Core Team: R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2013, URL <http://www.R-project.org/>.
58. Love MI, Huber W, Anders S: Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014, 15(12):550.
59. Warnes GR, Bolker B, Bonebakker L, Gentleman R, Huber W, Liaw A, Lumley T, Maechler M, Magnusson A, Moeller S, *et al*: gplots: Various R Programming Tools for Plotting Data. R package version 3.0.1 2016, <https://CRAN.R-project.org/package=gplots>

Table 4.1 The rank of expression of lignin biosynthetic genes in wild type samples

FPKM (fragments per kilobase of transcript per million mapped reads) shows the average expression of each gene from eight biological replicates. The rank was based on FPKM from high to low among all 20974 expressed genes.

ID	Gene name	FPKM	Rank
AT2G37040	<i>PAL1</i>	353.3	80
AT3G53260	<i>PAL2</i>	130.8	297
AT5G04230	<i>PAL3</i>	5.1	10313
AT3G10340	<i>PAL4</i>	157.0	231
AT2G30490	<i>C4H</i>	468.1	56
AT1G51680	<i>4CL1</i>	298.9	102
AT3G21240	<i>4CL2</i>	33.5	1683
AT1G65060	<i>4CL3</i>	0.5	18157
AT3G21230	<i>4CL4</i>	21.6	2839
AT5G48930	<i>HCT</i>	248.6	132
AT2G40890	<i>C3'H</i>	137.8	275
AT1G52760	<i>CSE</i>	189.3	179
AT4G34050	<i>CCoAOMT1</i>	1020.0	16
AT1G15950	<i>CCR1</i>	120.3	326
AT1G80820	<i>CCR2</i>	1.7	15157
AT4G36220	<i>F5H</i>	589.5	40
AT5G54160	<i>COMT1</i>	751.7	24
AT4G39330	<i>CAD9</i>	29.8	1954
AT3G19450	<i>CAD4</i>	40.9	1326
AT4G34230	<i>CAD5</i>	141.1	269
AT3G24503	<i>REF1</i>	20.8	2970
AT3G21560	<i>BRT1</i>	0.9	16966
AT2G22990	<i>SNG1</i>	46.4	1136
AT2G38080	<i>LAC4</i>	249.8	131
AT5G03260	<i>LAC11</i>	34.6	1623
AT5G60020	<i>LAC17</i>	74.7	601

Table 4.2 Concentrations of 18 phenylpropanoids from wild type and mutants.

Metabolites were quantified on LC/MS-MS. Mean and standard deviation (SD) were calculated. One-way ANOVA with Dunnett's test was applied to test each metabolite in mutant compared to wild type. p-value less than 0.001 is designed as ***, p-value between 0.001 and 0.01 is designed as **, and p-value between 0.01 and 0.05 is designed as *.

See Excel file Table4.2_MetabolitesConcentrations.xlsx.

Table 4.3 Gene ontology enrichment of gene clusters from the hierarchical clustering results.

AGI code of genes from each cluster was listed and used for enriched gene ontology analysis. The significantly enriched gene ontology terms (FDR < 0.1) were shown and included in Figure 4.13.

See Excel file Table4.3_ClusteringGOTerms.xlsx.

Table 4.4 Common DEGs and enriched gene ontology from subgroups of mutants.

The shared up or down regulated DEGs from each subgroup were listed with the log₂FC changes compared to wild type, adjusted *p*-values, and short description of the gene. The list of DEGs were used for gene ontology analysis and enriched terms were listed in the table.

See Excel file Table4.4_Subgroups_DEGsAndGOTerms.xlsx

Table 4.5 Primer sequences for genotyping the Arabidopsis mutants

Primer ID	Primer name	Sequences (5'-3')
CC4010	<i>pal1</i> LP	CCACCTTATTAAACCACCCTTC
CC4011	<i>pal1</i> RP	ATGTGTGGCTTGTTTCTTTTCG
CC2449	<i>pal1</i> BP	ATTTTGCCGATTTTCGGAAC
CC4061	<i>pal2</i> LP	CAATGGATCAAATCGAAGCA
CC4062	<i>pal2</i> RP	TATTCCGGCGTTCAAAAATC
CC2501	<i>pal2</i> BP	ATATTGACCATCATACTCATTGC
CC2396	<i>ref3-2</i> LP	TTCCGTATCATGTTCGATAG
CC2397	<i>ref3-2</i> RP	AATGTCAATTTCCCAAATC
CC2707	<i>ref3-3</i> LP	TCTTTAATCGCCGTCTTCGT
CC2708	<i>ref3-3</i> RP	TGGCACATTTCAATCCTTCA
CC3991	<i>4cl1</i> LP	ATCACCTCAGCAAAATCATG
CC3992	<i>4cl1</i> RP	CTGGATCAGCTCCAATAGCAG
CC4208	<i>4cl1</i> BP	GTGTGAGTAGTTCCCAGATAAG
CC5000	<i>cse-2</i> LP	AAAACACATCAAAACGATGCC
CC5001	<i>cse-2</i> RP	CTCTCCTTGAATCAGCGAGTG
CC2449	<i>cse-2</i> BP	ATTTTGCCGATTTTCGGAAC
CC3325	<i>cadC</i> LP	AGGTGAGGTGTTGGAAGTGG
CC3326	<i>cadC</i> RP	AAGCCTCTCAAACGCAATGT
CC3089	<i>cadC</i> BP	GCTTCCTATTATATCTTCCCAAATTACCAATACA
CC2461	<i>cadD</i> LP	TCACCGCAATCAAATAAAACC
CC2460	<i>cadD</i> RP	GACCCACATTGTTTGTTTAC
CC3089	<i>cadD</i> BP	GCTTCCTATTATATCTTCCCAAATTACCAATACA
CC1660	<i>fah1</i> LP	TGGTGTGTACATATATGGATGAAGAA
CC1661	<i>fah1</i> RP	TAGCAAGAGTGGTGAATATGTGAAGT
CC1863	<i>C4H:F5H</i>	AAGTGCCAGTGATTTTGATCG
CC1864	<i>C4H:F5H</i>	CGGACCATTTTGTCCACTTC
CC1742	<i>SmF5H</i>	CAAGGTCCTCCACAAGAAGC
CC1743	<i>SmF5H</i>	CAGTCGAAGCACTGGATGAA
CC1670	<i>ref8</i> LP	CGAGCTATCATGGAGGAGCATA
CC0247	<i>ref8</i> RP	CAACAAGAGCATGAGCAGCAG
CC1834	<i>med5a</i> LP	TTTTATGGGCCTCTTTCGTG
CC1776	<i>med5a</i> RP	TTGGCATTAGTGAGCAAGCA
CC2449	<i>med5a</i> BP	ATTTTGCCGATTTTCGGAAC
CC1446	<i>med5b</i> LP	ACGGGGCATTGATAGAAAAA
CC1554	<i>med5b</i> RP	AGGAGGGAATCGACAATGTG
CC2449	<i>med5a</i> BP	ATTTTGCCGATTTTCGGAAC
CC1612	<i>ref2</i> LP	GCGGTTCTCCTTTTCTTCT
CC1613	<i>ref2</i> RP	ACAAGATTGGACCGTATTTTGGGC

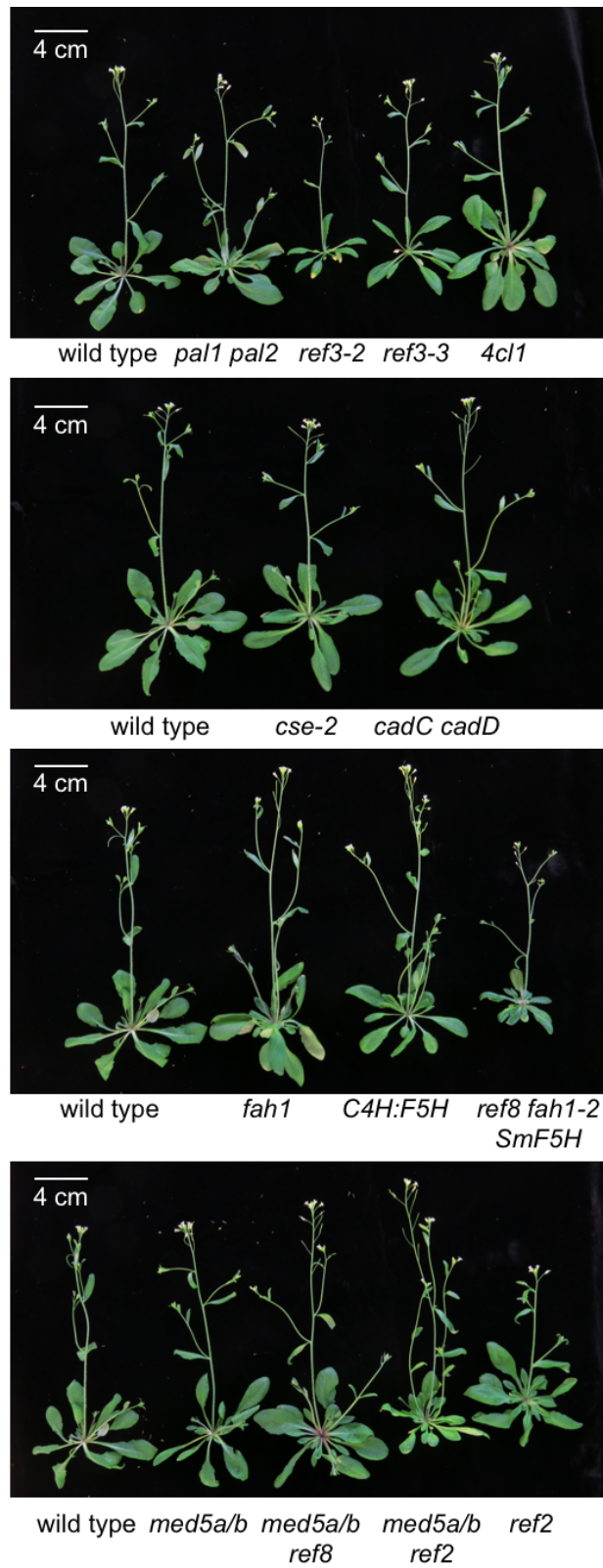


Figure 4.1 Growth phenotype of 28-day-old wild-type *Arabidopsis* and plants with perturbed lignification.

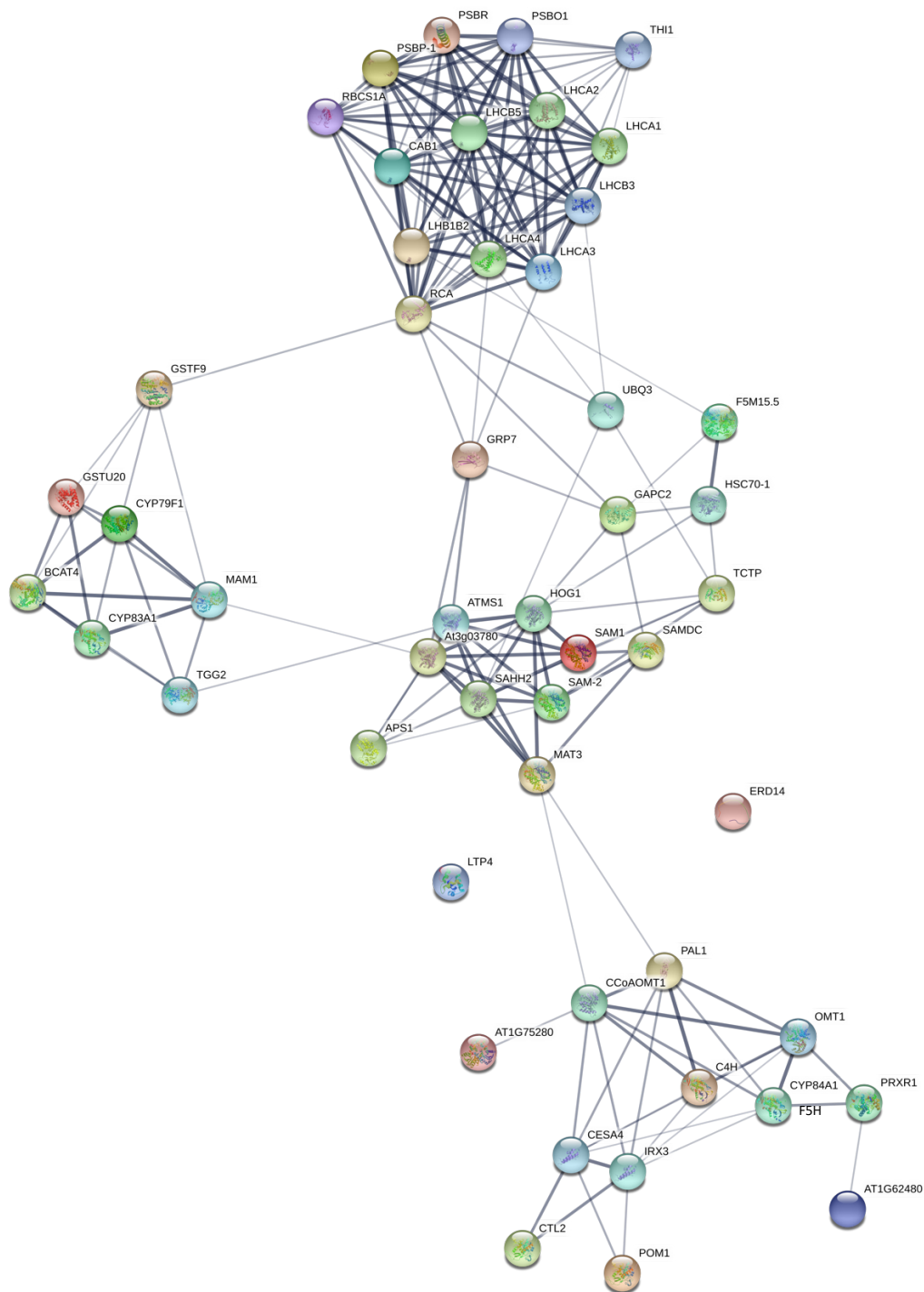


Figure 4.2 Association of the 50 genes with the highest expression.

The associates between the 50 genes with highest averaged expression across all 60 samples were analyzed in STRING. Association between two genes was determined based on co-expression, experimental/biochemical data, association in curated databases, and co-mention in PubMed abstracts. Lines indicate interaction between two genes with a confidence score higher than 0.4 and thickness of the line indicates the strength of data support.

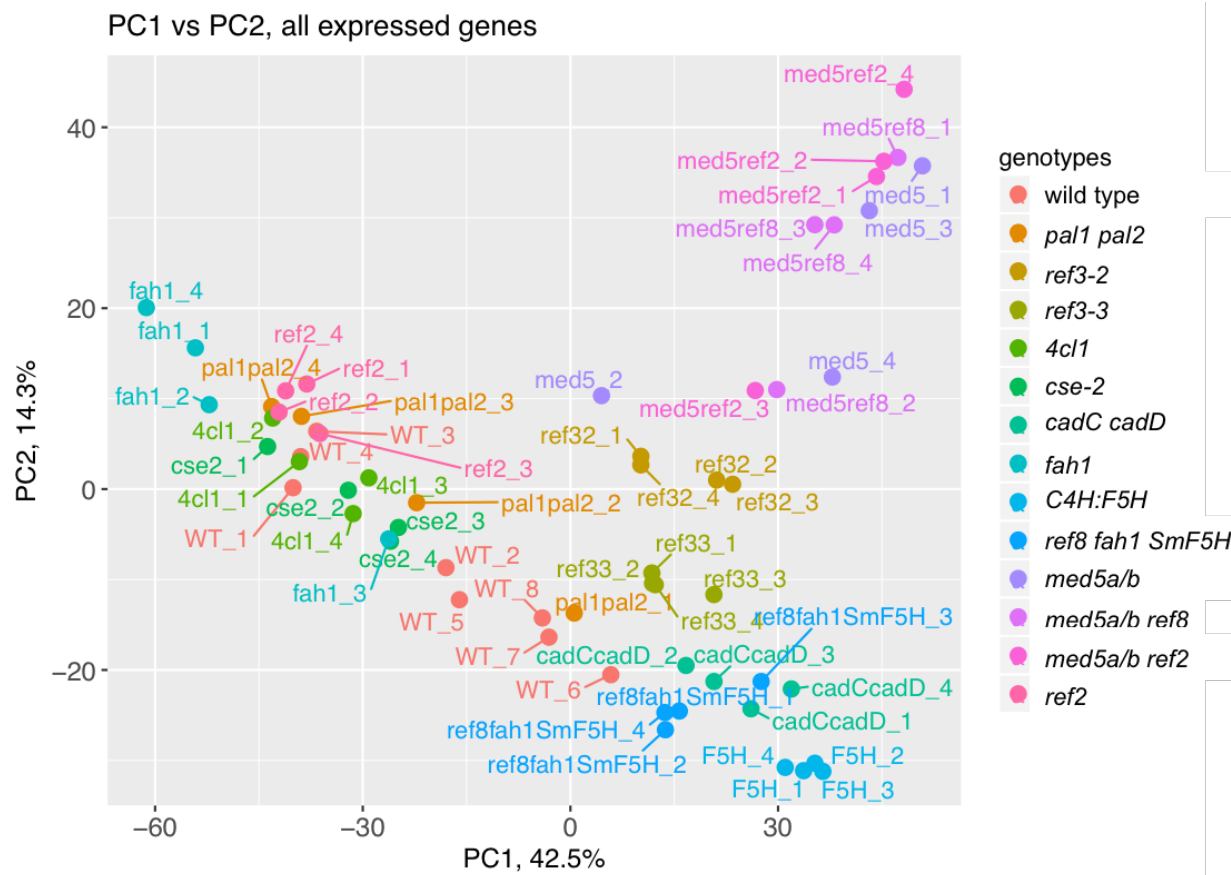


Figure 4.3 PCA analysis of gene expression profiles in wild type and the perturbed plants.

PCA analysis of expression of genes from individual samples in RNA-seq data. The values were determined with log transformed transcript counts. The first two principle components were plotted here. Percentage on X and Y axis showed the variance explained by PC1 and PC2, respectively. On the figure, wild type was labeled as WT, *ref3-2* and *ref3-3* as ref32 and ref33, *cse-2* as cse2, *C4H:F5H* as F5H, *med5a/b* as med5.

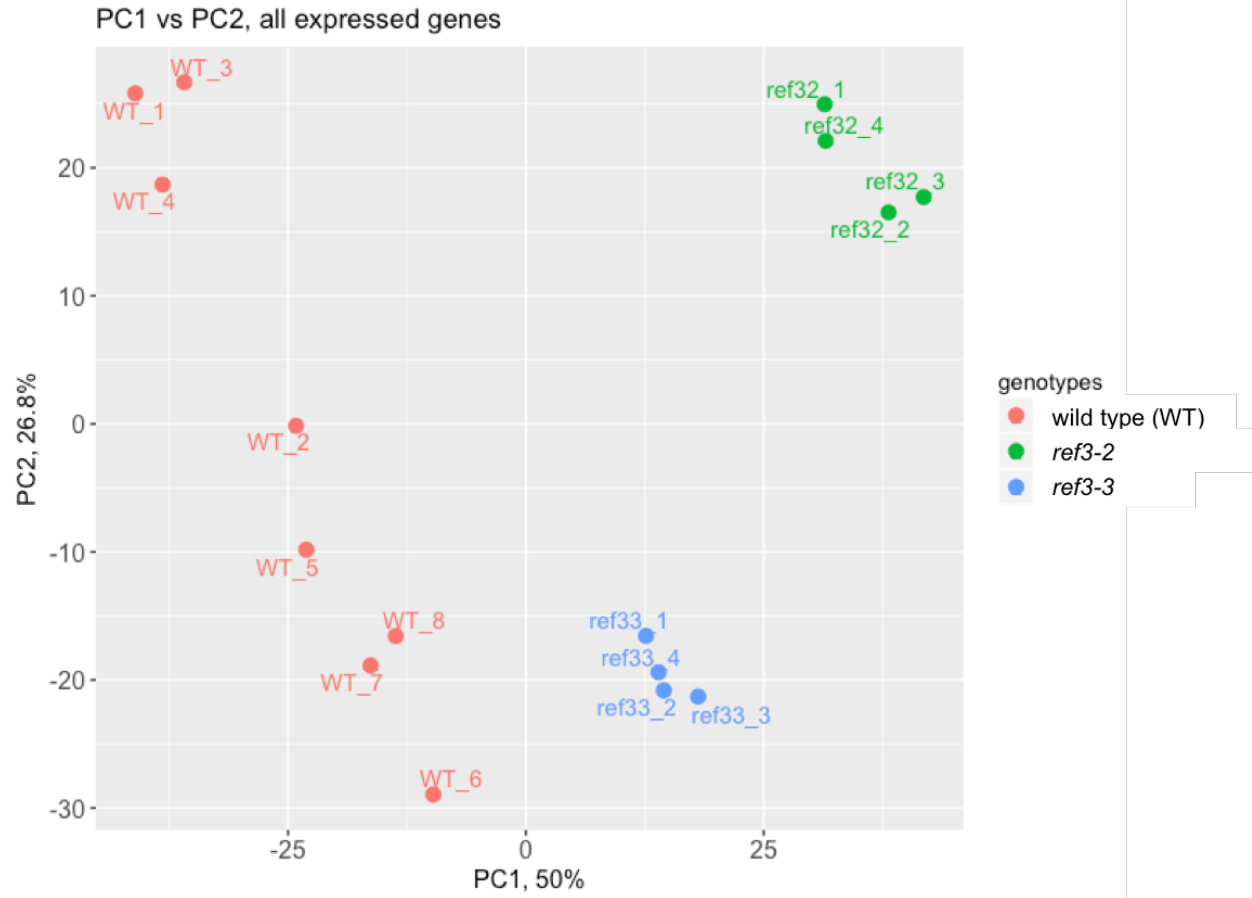


Figure 4.4 PCA analysis of expressed genes in RNA-seq data from individual samples of wild type, *ref3-2*, and *ref3-3*.

The values were determined with log transformed transcript counts. The first two principle components were plotted here. Percentage on X and Y axis showed the variance explained by PC1 and PC2, respectively. On the figure, wild type was labeled as WT, *ref3-2* and *ref3-3* as *ref32* and *ref33*.

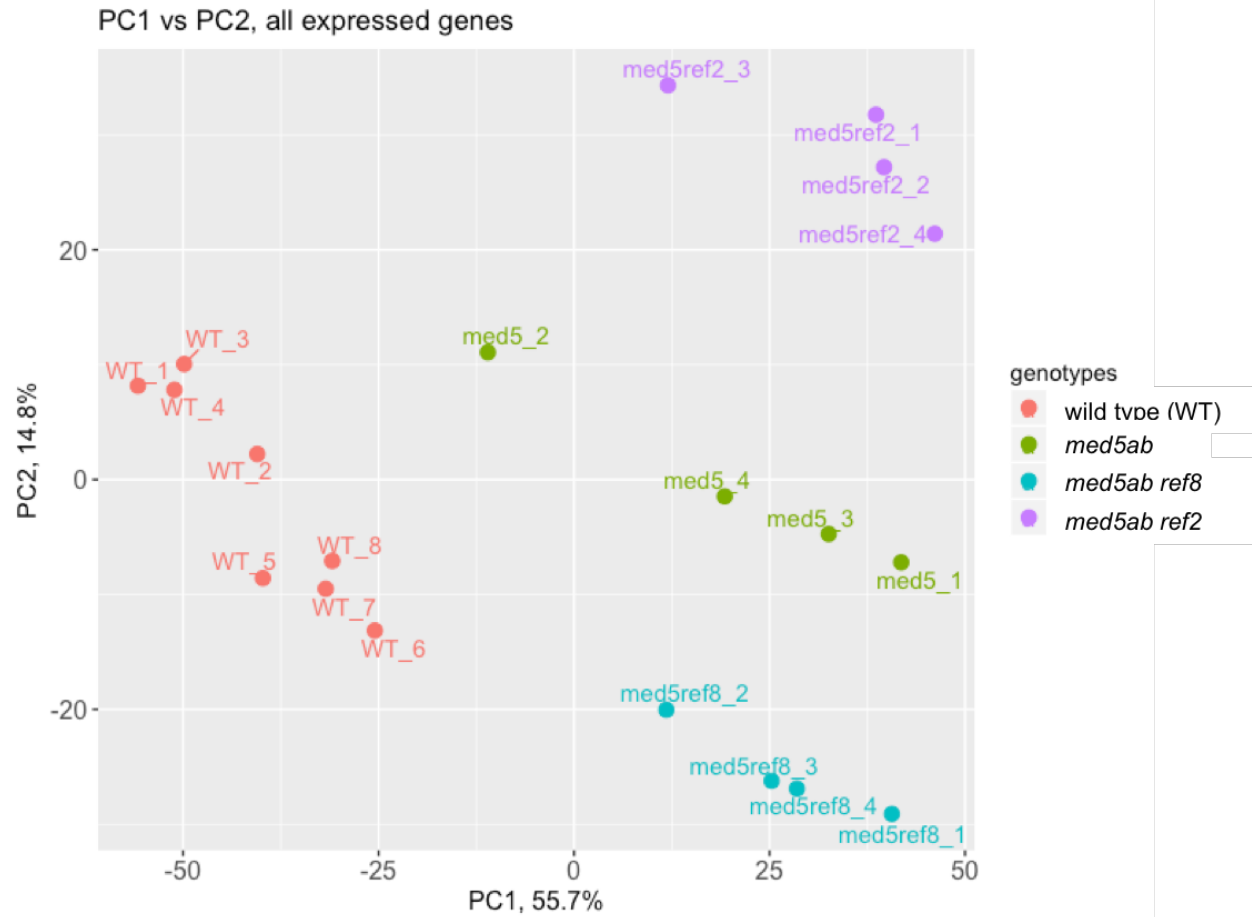


Figure 4.5 PCA analysis of expressed genes in RNA-seq data from individual samples of wild type, *med5a/b*, *med5a/b ref8*, and *med5ab ref2*.

The values were determined with log transformed transcript counts. The first two principle components were plotted here. Percentage on X and Y axis showed the variance explained by PC1 and PC2, respectively. On the figure, wild type was labeled as WT, *med5a/b* as med5.

Euclidean distance between individual samples were calculated based on transcription of all expressed genes. Blue gradient indicates the distance between samples as dark blue means close while light blue means distant. On the figure, wild type was labeled as WT, *ref3-2* and *ref3-3* as ref32 and ref33.

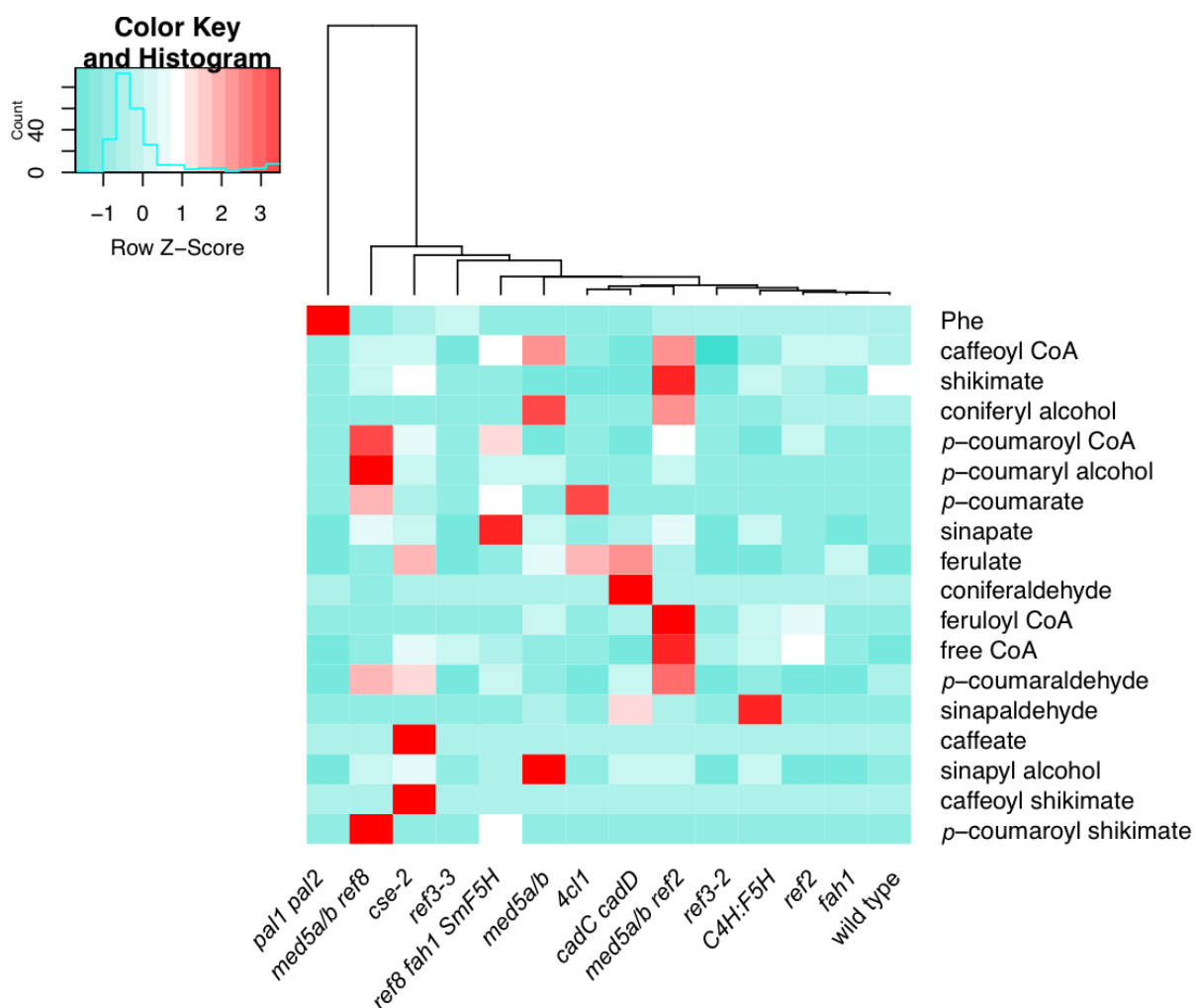


Figure 4.7 Metabolic profiles from wild type and the perturbed plants.

Concentrations of 18 phenylpropanoid metabolites were measured in the basal stem tissue on LC/MS-MS. Concentration means were scaled *per* metabolite across all genotypes to generate heatmap. Wild type and mutants were clustered based on metabolite profiles using ward.D2 method.

Gene	<i>pal1</i> <i>pal2</i>	<i>ref3-2</i>	<i>ref3-3</i>	<i>4cl1</i>	<i>cse-2</i>	<i>cadC</i> <i>cadD</i>	<i>fah1</i>	<i>C4H:</i> <i>F5H</i>	<i>ref8</i> <i>fah1</i> <i>SmF5H</i>	<i>med5</i> <i>a/b</i>	<i>med5</i> <i>a/b</i> <i>ref8</i>	<i>med5</i> <i>a/b</i> <i>ref2</i>	<i>ref2</i>
<i>PAL1</i>	-3.7	0.7	0.5	-0.2	-0.7	0.4	-0.8	0.2	0.0	0.2	0.6	0.0	-0.9
<i>PAL2</i>	-2.6	1.0	0.6	0.4	-0.9	0.2	-0.7	-0.2	-0.3	0.6	0.9	0.3	-0.8
<i>PAL3</i>	0.3	0.4	0.6	0.1	-0.2	0.2	0.1	-0.1	-0.1	0.3	0.7	1.1	0.5
<i>PAL4</i>	0.1	1.0	0.8	0.1	-1.0	0.2	-1.0	0.2	0.3	1.0	1.2	0.6	-1.2
<i>C4H</i>	0.0	0.9	0.6	0.1	-0.9	0.1	-0.7	-0.1	-0.1	0.6	0.8	0.3	-0.9
<i>4CL1</i>	-0.1	0.6	0.5	-5.1	-0.9	0.2	-0.6	0.1	0.0	0.3	0.6	0.3	-0.5
<i>4CL2</i>	0.5	1.8	1.3	0.4	-0.6	1.3	-0.8	1.1	1.3	1.3	2.0	1.2	-0.8
<i>4CL3</i>	0.9	-0.2	-0.5	0.3	0.4	-0.5	1.2	-1.3	-1.0	-0.6	-0.1	-0.3	0.2
<i>4CL4</i>	-0.2	0.8	0.6	-0.2	-0.5	0.8	-1.0	0.6	0.8	0.5	0.8	0.4	-0.9
<i>HCT</i>	-0.1	0.4	0.3	-0.2	-0.7	0.3	-0.5	-0.2	-0.1	0.1	0.2	0.0	-0.5
<i>C3'H</i>	0.2	0.7	0.4	0.2	-0.7	0.1	-0.3	-0.4	-0.3	0.3	0.8	0.0	-0.5
<i>CSE</i>	-0.2	0.3	0.3	-0.2	-4.4	0.2	-0.6	0.1	0.3	0.1	0.4	-0.1	-0.6
<i>CCoAOMT1</i>	0.0	0.5	0.3	0.0	-0.5	0.2	-0.6	0.0	0.0	0.1	0.4	-0.1	-0.6
<i>CCR1</i>	-0.1	0.3	0.0	0.0	-0.2	0.2	-0.3	0.2	0.1	0.0	0.2	0.0	-0.3
<i>CCR2</i>	1.1	4.1	1.7	1.2	-0.2	1.6	-0.7	2.7	3.2	2.2	3.5	2.2	-0.5
<i>F5H</i>	-0.4	0.5	0.6	-0.8	-0.8	0.5	-3.9	5.9	-2.2	0.5	0.7	0.6	-1.1
<i>COMT1</i>	-0.2	0.6	0.4	-0.3	-0.6	0.3	-0.9	0.3	0.5	0.1	0.3	0.1	-0.8
<i>CAD4</i>	0.1	0.9	0.4	0.1	-0.7	-2.8	-0.6	0.2	0.1	0.5	0.5	0.3	-0.3
<i>CAD5</i>	-0.3	0.7	0.2	-0.2	-0.6	-13.6	-0.6	0.6	0.4	0.4	0.6	0.1	-0.5
<i>REF1</i>	-0.3	0.0	0.0	0.1	-0.1	0.1	-0.4	0.2	-0.6	0.6	0.3	0.9	0.5
<i>BRT1</i>	0.0	1.1	-0.3	0.0	0.3	0.4	0.0	0.7	0.2	1.7	0.8	2.8	0.4
<i>SNG1</i>	0.6	0.4	0.6	0.0	0.2	0.0	0.1	-0.6	0.5	-0.3	0.3	-0.5	0.5
<i>LAC4</i>	-0.3	0.0	-0.1	-0.1	-0.4	0.1	-0.4	-0.2	-0.1	-0.5	-0.1	-0.9	-0.2
<i>LAC11</i>	-0.1	0.7	0.4	0.0	0.2	0.6	-0.8	1.8	0.8	-0.5	0.8	-1.2	-0.8
<i>LAC17</i>	-0.4	-0.3	-0.4	-0.4	0.0	0.1	-0.4	0.1	0.4	-0.7	-0.5	-1.6	0.1
<i>PRX64</i>	0.2	-1.0	-0.6	0.5	0.2	-1.2	0.1	-2.1	-1.2	-0.8	-0.7	-1.6	0.2
<i>PRX72</i>	1.3	-2.5	-2.3	1.1	1.0	-0.7	2.8	3.7	0.4	2.1	-1.9	3.2	0.2

Figure 4.8 Expression of lignin biosynthetic genes in the RNA-seq data.

Values are log₂FC of expression in mutant compared to wild type. Bold indicates the value is significant (*p*-value cut-off of 0.05). Green background indicates genes mutated in the line.

Gene	<i>pal1</i> <i>pal2</i>	<i>ref3-2</i>	<i>ref3-3</i>	<i>4cl1</i>	<i>cse-2</i>	<i>cadC</i> <i>cadD</i>	<i>fah1</i>	<i>C4H:</i> <i>F5H</i>	<i>ref8</i> <i>fah1</i> <i>SmF5H</i>	<i>med5</i> <i>a/b</i>	<i>med5</i> <i>a/b</i> <i>ref8</i>	<i>med5</i> <i>a/b</i> <i>ref2</i>	<i>ref2</i>
<i>DAHS1</i>	0.3	1.3	0.9	0.3	-0.6	-0.1	-0.1	-0.6	-0.4	0.8	1.3	1.0	0.0
<i>DAHS2</i>	0.2	0.5	0.2	0.1	-0.1	0.3	0.2	0.4	0.4	0.3	0.2	0.6	0.5
<i>DAHS3</i>	0.0	0.8	0.4	0.1	-0.7	0.1	-0.7	0.0	-0.1	0.4	0.8	0.2	-0.8
<i>DHQS</i>	-0.1	0.4	0.2	0.1	-0.3	0.3	-0.3	0.3	0.1	0.4	0.6	0.4	-0.2
<i>DHQS/SD</i>	-0.1	0.2	0.2	-0.1	-0.4	0.1	-0.4	0.0	0.0	0.2	0.6	0.0	-0.4
<i>SK1</i>	-0.3	-0.1	0.1	-0.2	-0.3	0.2	-0.4	0.1	-0.1	0.1	0.0	0.0	-0.3
<i>SK2</i>	0.1	0.4	0.3	0.0	-0.3	0.1	-0.2	-0.1	-0.1	0.2	0.4	0.2	0.0
<i>EPSPS</i>	0.0	0.4	0.2	0.2	-0.7	0.0	-0.3	0.0	-0.4	0.5	0.6	0.3	-0.5
<i>CS</i>	0.0	0.1	0.0	0.0	-0.2	0.0	-0.1	-0.1	-0.1	0.1	0.3	0.2	0.1
<i>CM1</i>	-0.1	0.1	0.2	0.0	-0.6	0.1	-0.6	-0.1	-0.5	0.3	0.6	0.0	-0.7
<i>CM2</i>	0.0	-0.1	0.1	0.0	-0.1	-0.2	-0.1	-0.1	-0.2	-0.1	0.0	0.0	0.0
<i>CM3</i>	0.1	0.3	0.1	0.1	0.1	0.0	-0.2	-0.3	0.1	-0.2	0.1	-0.3	0.1
<i>PPA-AT</i>	-0.1	0.3	0.2	-0.1	-0.5	0.0	-0.5	-0.3	-0.3	0.0	0.4	-0.3	-0.6
<i>ADT1</i>	0.0	0.0	-0.3	-0.1	-0.1	0.0	0.0	0.3	-0.2	-0.3	-0.4	-0.4	0.0
<i>ADT2</i>	0.2	0.2	0.1	0.1	0.1	0.3	0.3	0.2	0.0	0.4	0.4	0.6	0.3
<i>ADT3/PDT1</i>	0.3	0.2	0.2	0.0	-0.2	-0.2	0.0	-0.3	-0.2	-0.2	0.0	0.7	0.6
<i>ADT4</i>	0.1	0.8	0.5	-0.3	-0.6	0.4	-0.4	0.0	0.3	0.8	1.2	0.6	-0.6
<i>ADT5</i>	0.0	1.0	0.7	-0.1	-0.4	0.4	-0.8	0.3	1.0	0.7	1.2	0.4	-0.8
<i>ADT6</i>	0.3	0.7	0.6	0.0	-0.4	0.3	-0.1	0.4	-0.1	0.5	0.5	0.6	0.1

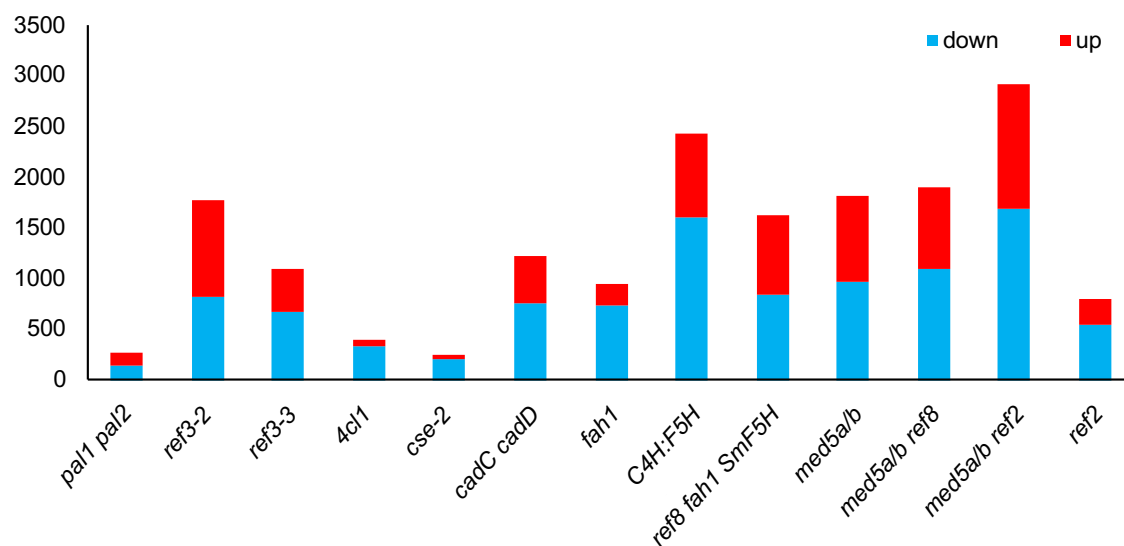
Figure 4.9 Expression of genes involved in shikimate biosynthesis in the RNA-seq data.

Values are log₂FC of expression in mutant compared to wild type. Bold indicates the value is significant (*p*-value cut-off of 0.05). Green background indicates genes mutated in the line.

Gene	<i>pal1</i> <i>pal2</i>	<i>ref3-2</i>	<i>ref3-3</i>	<i>4cl1</i>	<i>cse-2</i>	<i>cadC</i> <i>cadD</i>	<i>fah1</i>	<i>C4H:</i> <i>F5H</i>	<i>ref8</i> <i>fah1</i> <i>SmF5H</i>	<i>med5</i> <i>a/b</i>	<i>med5</i> <i>a/b</i> <i>ref8</i>	<i>med5</i> <i>a/b</i> <i>ref2</i>	<i>ref2</i>
<i>KNAT7</i>	-0.5	-0.3	-0.2	-0.4	-0.3	0.2	-0.7	0.3	0.3	-0.1	-0.3	-0.4	-0.5
<i>MYB103</i>	-0.3	-0.6	-0.2	0.1	-0.3	-0.1	-0.5	-0.3	-0.6	-0.5	-0.8	-1.1	-0.4
<i>MYB113</i>	-0.3	2.8	1.1	-0.3	0.2	0.1	-1.0	1.2	0.8	3.1	3.3	3.2	-0.4
<i>MYB114</i>	-0.3	1.6	0.6	-0.5	-0.4	-0.6	-1.1	0.7	-0.8	0.4	-0.1	1.2	-0.6
<i>MYB12</i>	0.4	-0.8	-0.1	-0.6	-0.4	-0.3	0.2	-1.0	-1.0	-1.0	-2.5	-0.9	-0.1
<i>MYB20</i>	-0.2	0.2	0.2	-0.8	-0.7	0.2	-0.9	-0.4	0.3	-0.1	0.4	-0.6	-0.7
<i>MYB4</i>	-0.5	-1.3	-0.6	-0.5	0.4	0.2	-0.2	-0.3	-0.1	-1.5	-1.6	-1.2	0.2
<i>MYB42</i>	-0.2	1.3	0.7	-0.1	-0.4	0.6	-0.8	0.9	0.9	0.6	1.0	0.7	-0.8
<i>MYB43</i>	0.0	0.1	0.4	-0.1	-0.8	-0.2	-0.5	-0.2	-0.3	0.5	0.5	0.1	-0.7
<i>MYB46</i>	-0.4	-0.5	-0.6	0.4	0.0	-0.2	-0.3	-0.5	-0.1	-0.5	-0.5	-1.0	0.0
<i>MYB52</i>	-0.4	0.0	0.0	-0.7	-0.4	0.3	-0.6	0.3	0.6	-0.3	-0.1	-0.5	-0.4
<i>MYB54</i>	-0.9	-0.8	-0.8	-0.3	-0.3	-0.1	-0.4	-0.1	0.1	-1.2	-1.6	-1.6	-0.6
<i>MYB58</i>	-0.1	-0.1	0.0	-0.1	-1.0	-0.7	-0.3	-1.4	-1.4	-0.2	-0.1	-0.7	-0.6
<i>MYB63</i>	-0.2	-0.1	-0.1	-0.4	-0.9	0.2	-0.7	-0.3	-0.7	-0.3	0.2	-0.5	-0.7
<i>MYB69</i>	-0.6	-0.1	0.5	-1.0	-0.3	0.7	-1.1	0.8	0.8	-0.2	0.0	-0.3	-0.7
<i>MYB75</i>	0.5	1.6	-0.5	-0.5	-0.6	-1.2	0.0	0.3	-0.3	-2.0	-1.9	-1.7	-0.4
<i>MYB83</i>	-0.2	-0.5	-0.7	0.5	0.2	-0.4	-0.2	0.3	0.4	-0.8	-0.8	-1.0	0.0
<i>MYB85</i>	-0.4	0.4	0.5	-0.6	-0.5	0.5	-0.8	0.3	0.5	-0.1	0.3	0.0	-0.8
<i>MYB90</i>	-1.1	1.1	-0.3	-0.6	-1.6	0.3	-2.5	2.0	-1.0	1.4	0.4	3.9	-2.2
<i>NST1</i>	-0.3	-0.8	-0.4	-0.2	0.0	-0.2	-0.4	-0.9	-0.3	-0.3	-0.5	-0.7	-0.1
<i>NST2</i>	-0.3	0.0	0.2	-0.3	-0.4	0.3	-0.7	0.5	0.3	0.1	-0.3	0.0	-0.2
<i>SND1</i>	-0.5	-0.2	0.1	-0.7	-0.4	0.5	-0.9	0.4	0.4	-0.3	0.1	-0.5	-0.8
<i>SND2</i>	-0.6	-0.6	-0.3	-0.7	-0.4	-0.1	-0.8	-0.3	0.0	-0.6	-0.4	-1.1	-0.7
<i>SND3</i>	-0.3	-0.1	0.1	-0.4	-0.3	0.3	-0.7	0.5	0.5	0.0	0.1	-0.3	-0.6
<i>VND6</i>	0.0	0.8	0.1	0.4	0.2	0.0	-0.2	0.7	0.4	-0.1	0.1	0.2	0.1
<i>VND7</i>	-0.2	0.1	0.1	-0.6	0.0	0.2	-0.5	-0.4	-0.1	-0.5	0.1	-1.7	-0.3
<i>KFB01</i>	-0.3	0.1	0.0	-0.5	0.1	0.7	-0.7	-0.5	0.3	0.3	0.6	0.5	0.2
<i>KFB20</i>	0.4	0.2	0.6	-1.9	-0.2	1.2	-2.4	0.9	2.1	0.2	0.7	0.9	-1.9
<i>KFB39</i>	-2.2	-4.6	-2.8	-1.4	1.4	0.0	-0.5	-2.1	0.5	-5.1	-4.9	-6.3	-0.4
<i>KFB50</i>	-0.1	-0.5	-0.4	-0.8	0.1	0.1	-0.8	-0.4	0.7	-1.2	-0.8	-0.9	-0.3
<i>Med5A</i>	0.2	-0.4	0.0	0.3	-0.1	-0.3	0.1	-0.7	-0.7	-1.4	-1.3	-1.7	0.3
<i>Med5B</i>	0.1	0.2	0.2	-0.1	0.0	0.3	0.0	0.3	0.1	-1.4	-1.3	-1.2	0.1
<i>REF2</i>	0.1	0.2	0.3	0.0	-0.3	0.2	-0.4	0.0	0.3	-0.3	-0.5	-1.6	-2.0
<i>REF5</i>	0.1	0.3	0.3	0.4	0.3	0.1	-0.3	-0.7	0.2	1.1	0.8	0.1	0.2

Figure 4.10 Expression of genes involved in regulating phenylpropanoid metabolism.

Values are log₂FC of expression in mutant compared to wild type. Bold indicates the value is significant (*p*-value cut-off of 0.05). Green background indicates genes mutated in the line.



	<i>pal1 pal2</i>	<i>ref3-2</i>	<i>ref3-3</i>	<i>4cl1</i>	<i>cse-2</i>	<i>cadC cadD</i>	<i>fah1</i>	<i>C4H: F5H</i>	<i>ref8 fah1 SmF5H</i>	<i>med5 a/b</i>	<i>med5 a/b ref8</i>	<i>med5 a/b ref2</i>	<i>ref2</i>
up	127	959	423	54	54	453	213	831	788	840	811	1232	262
down	130	816	679	330	198	758	742	1597	842	972	1094	1686	533
total	257	1775	1102	384	252	1211	955	2428	1630	1812	1905	2918	795

Figure 4.11 Numbers of substantial DEGs in each perturbed plant compared to wild type.

Expression of genes were compared to wild type and genes with \log_2FC higher than 1 (up) or lower than -1 (down) at adjusted p -value cut-off of 0.01 were counted.

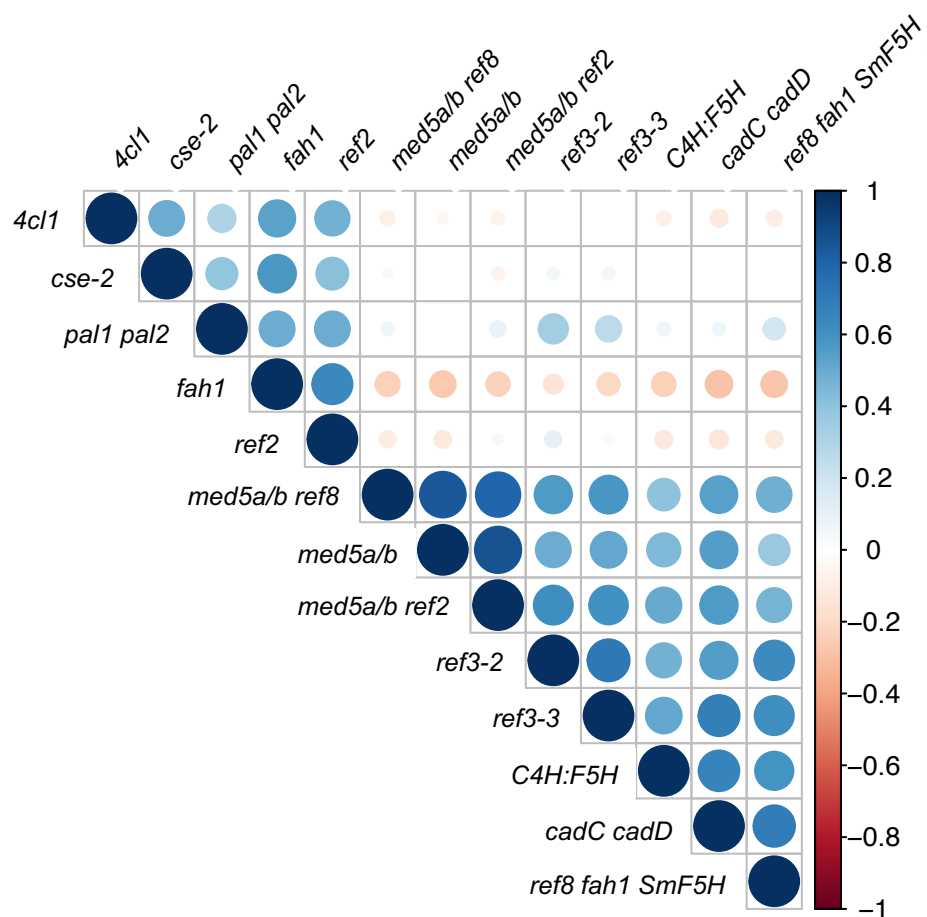


Figure 4.12 Pearson's correlation (r) of gene expression profiles.

Values were obtained based on \log_2FC of 5581 genes that were up or down regulated more than 2 fold in at least one mutant. Size of circle represents the absolute values of r and color code represents the value of r indicated as the scale on the right. Blank wells mean that the correlation is not significant based on p -value cut-off of 0.01.

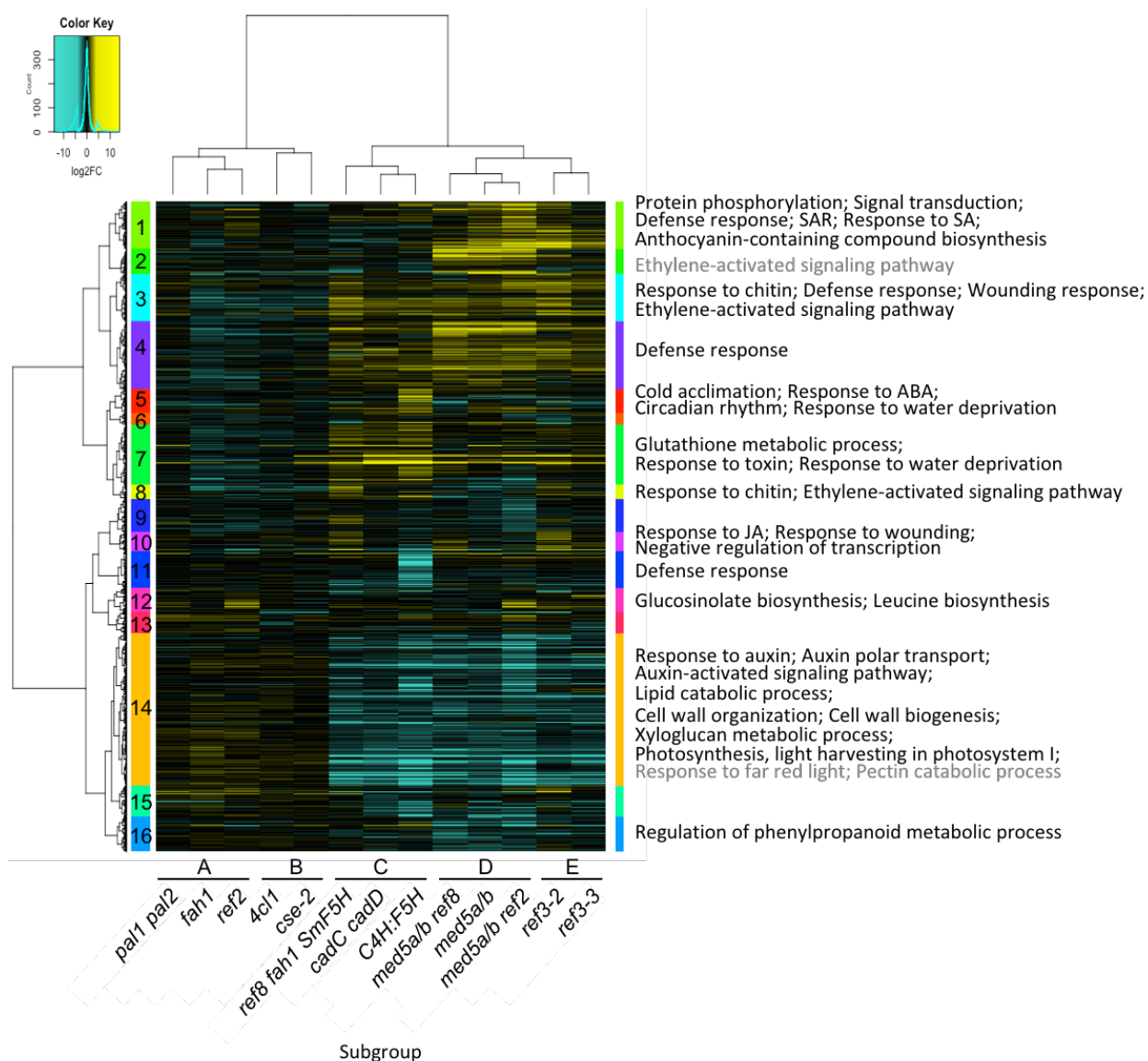


Figure 4.13 Hierarchical clustering analysis of substantial DEGs from 13 mutants compared with wild type.

Log₂FC values for 5581 DEGs with at least 2-fold changes in at least one mutant were used to generate the heat map and ward.D2 method was used for clustering. Gene ontology terms of biological processes for 16 clusters are shown on the right. Black color indicates statistically significant (FDR < 0.01) and grey color indicates FDR between 0.01 and 0.1.

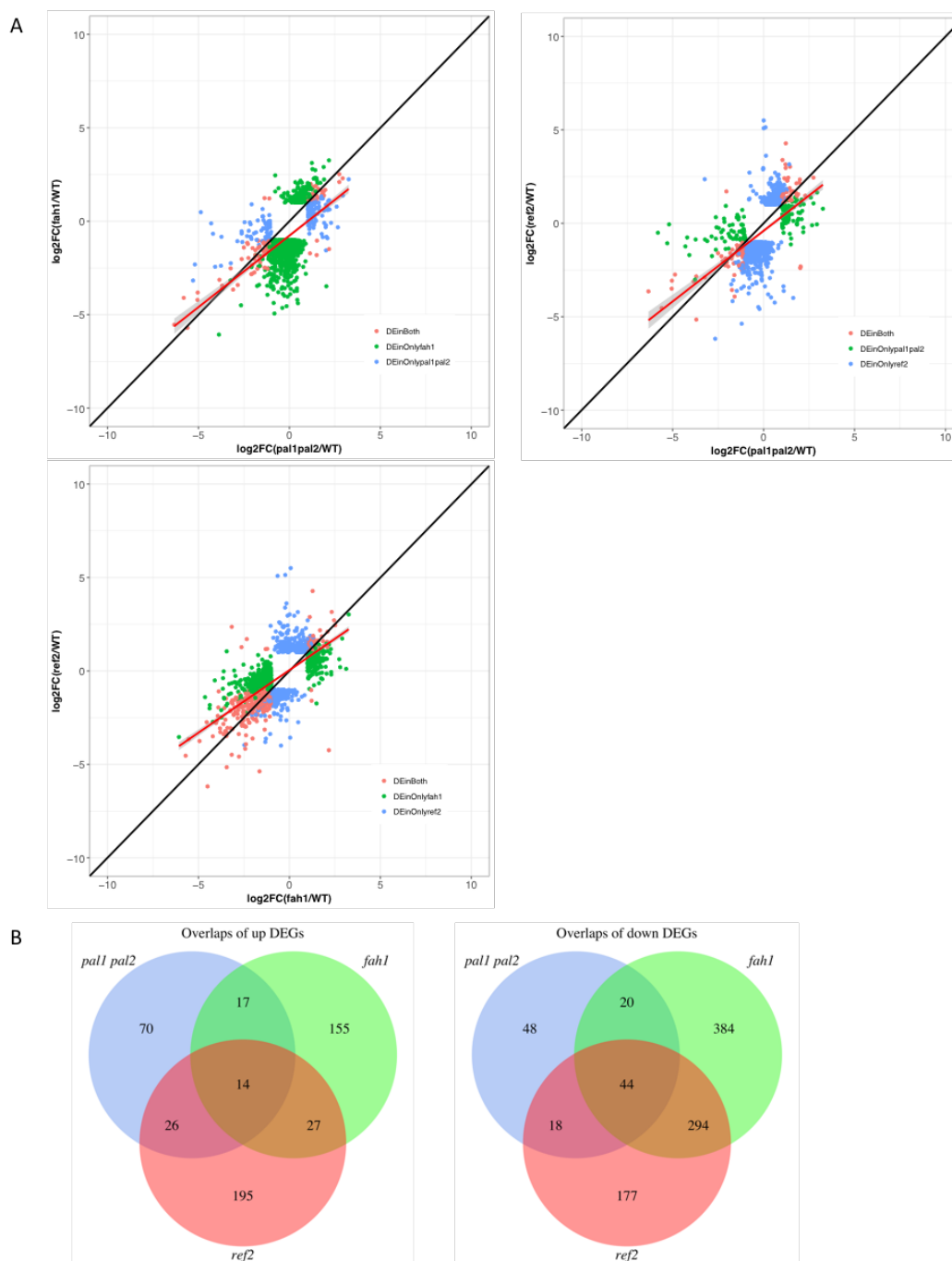
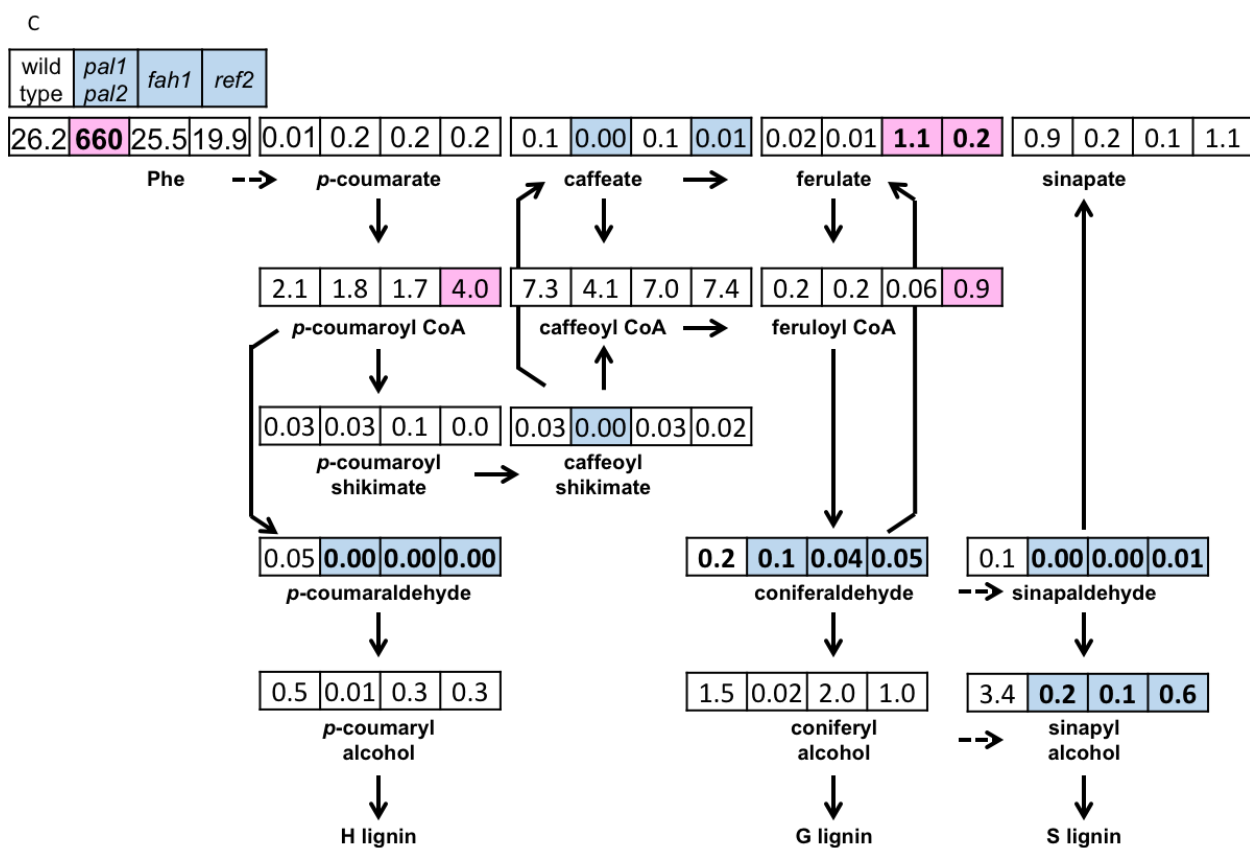


Figure 4.14 Comparison of metabolites and transcripts in wild type, *pal1 pal2*, *fah1*, and *ref2*.

(A) \log_2FC of substantial DEGs in *pal1 pal2*, *fah1* and *ref2*. (B) Venn diagram of up or down regulated genes in each mutant compared to wild type. (C) Concentrations (in the unit of nmol g FW^{-1}) of phenylpropanoids in wild type, *pal1 pal2*, *fah1*, and *ref2*. Numbers represent means of absolute concentrations of metabolites. One-way ANOVA with Dunnett's test was applied to compare metabolite concentration in each mutant to wild type. Background color indicates the value is significantly higher (pink) or lower (blue) compared to wild type at adjusted p -value of 0.05. Bold indicates the value is significant at adjusted p -value of 0.01.

Figure 4.14 continued



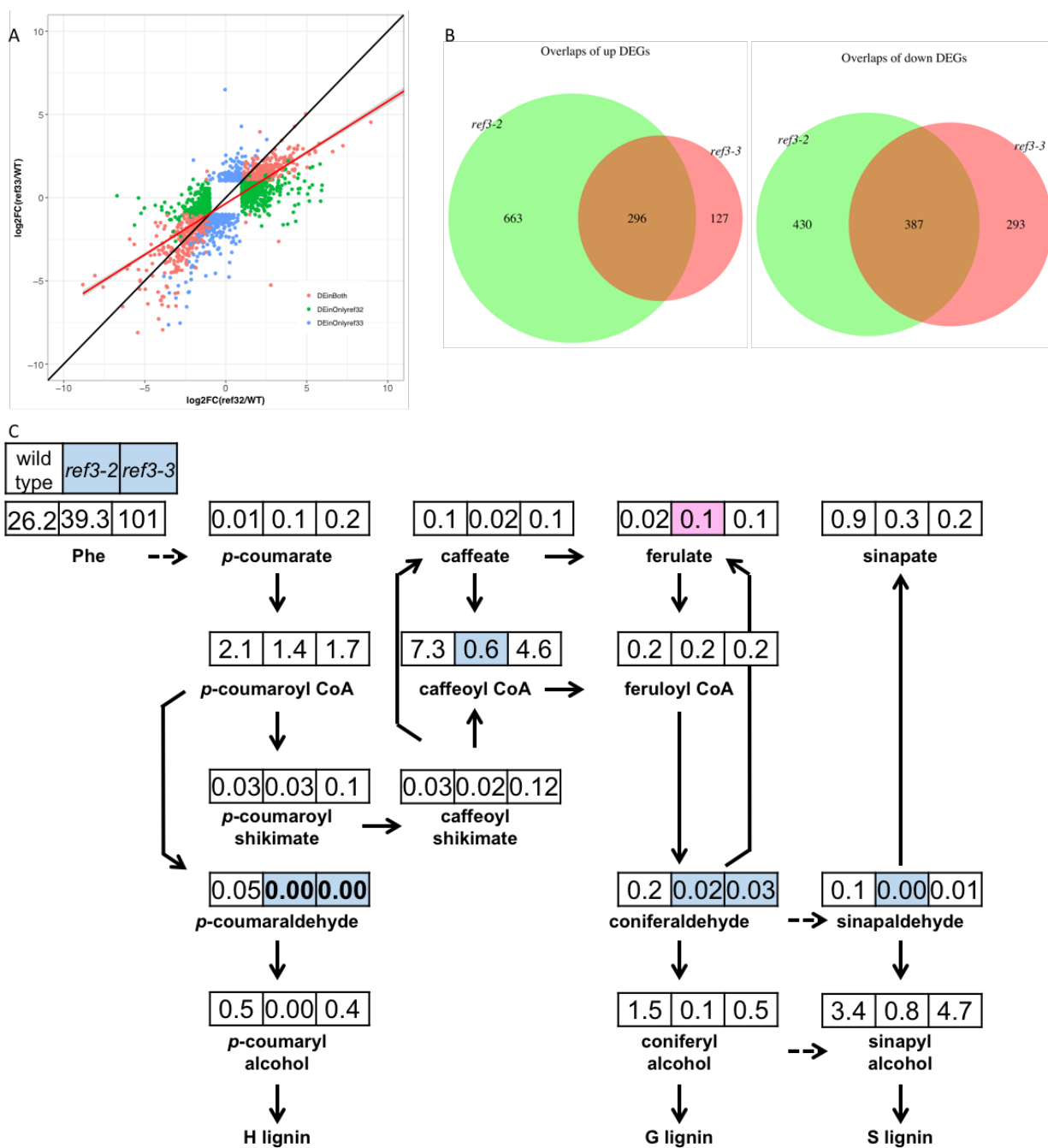


Figure 4.15 Comparison of metabolites and transcripts in wild type, *ref3-2*, and *ref3-3*.

(A) \log_2FC of substantial DEGs in *ref3-2*, and *ref3-3*. (B) Venn diagram of up or down regulated genes in each mutant compared to wild type. (C) Concentrations (in the unit of nmol g FW⁻¹) of phenylpropanoids in wild type, *ref3-2*, and *ref3-3*. Numbers represent means of absolute concentrations of metabolites. One-way ANOVA with Dunnett's test was applied to compare metabolite concentration in each mutant to wild type. Background color indicates the value is significantly higher (pink) or lower (blue) compared to wild type at adjusted *p*-value of 0.05. Bold indicates the value is significant at adjusted *p*-value of 0.01.

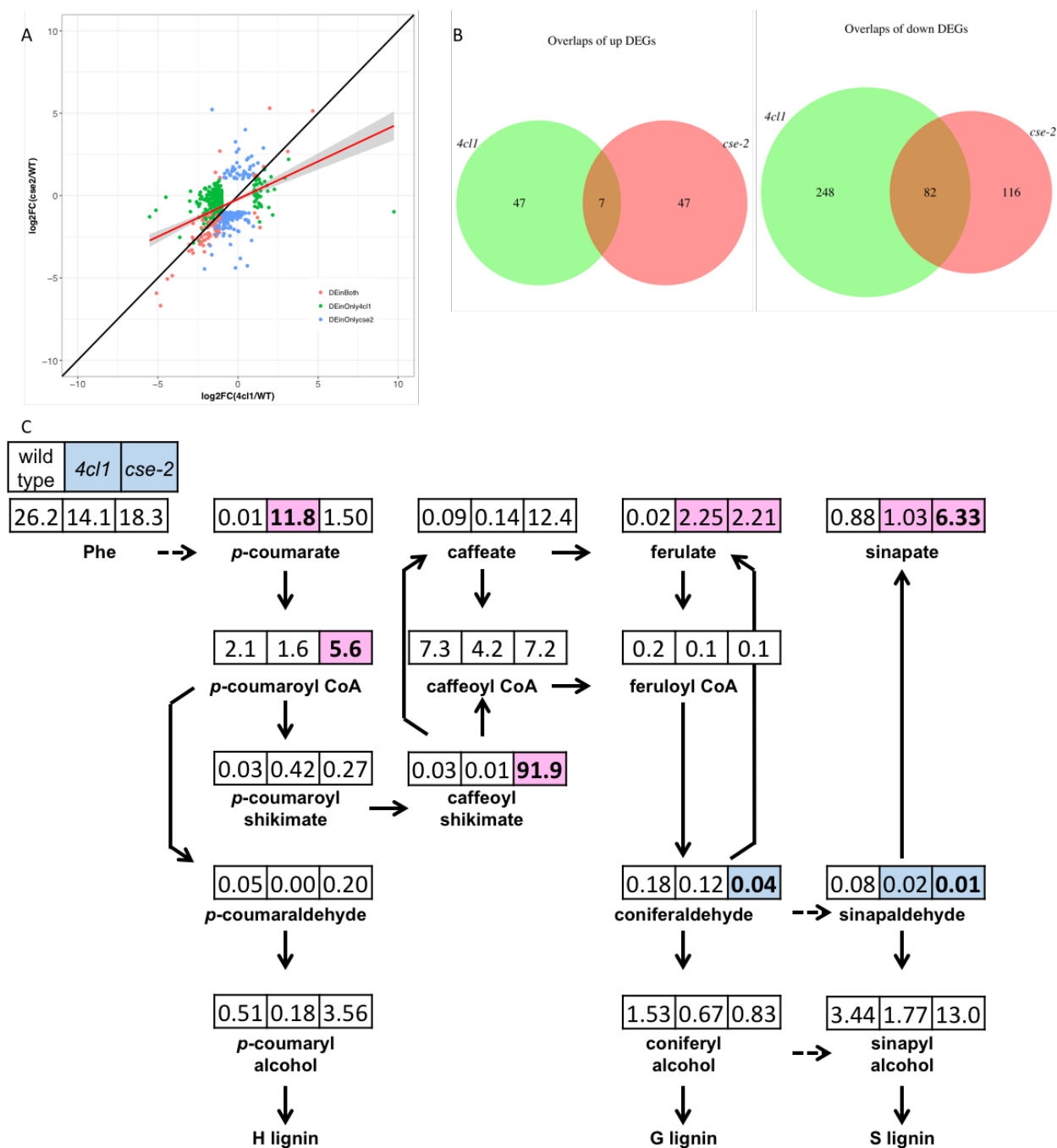


Figure 4.16 Comparison of metabolites and transcripts in wild type, *4cl1*, and *cse-2*.

(A) \log_2FC of substantial DEGs in *fah1* and *ref2*. (B) Venn diagram of up or down regulated genes in each mutant compared to wild type. (C) Concentrations (in the unit of nmol g FW^{-1}) of phenylpropanoids in wild type, *4cl1*, and *cse-2*. Numbers represent means of absolute concentrations of metabolites. One-way ANOVA with Dunnett's test was applied to compare metabolite concentration in each mutant to wild type. Background color indicates the value is significantly higher (pink) or lower (blue) compared to wild type at adjusted p -value of 0.05. Bold indicates the value is significant at adjusted p -value of 0.01.

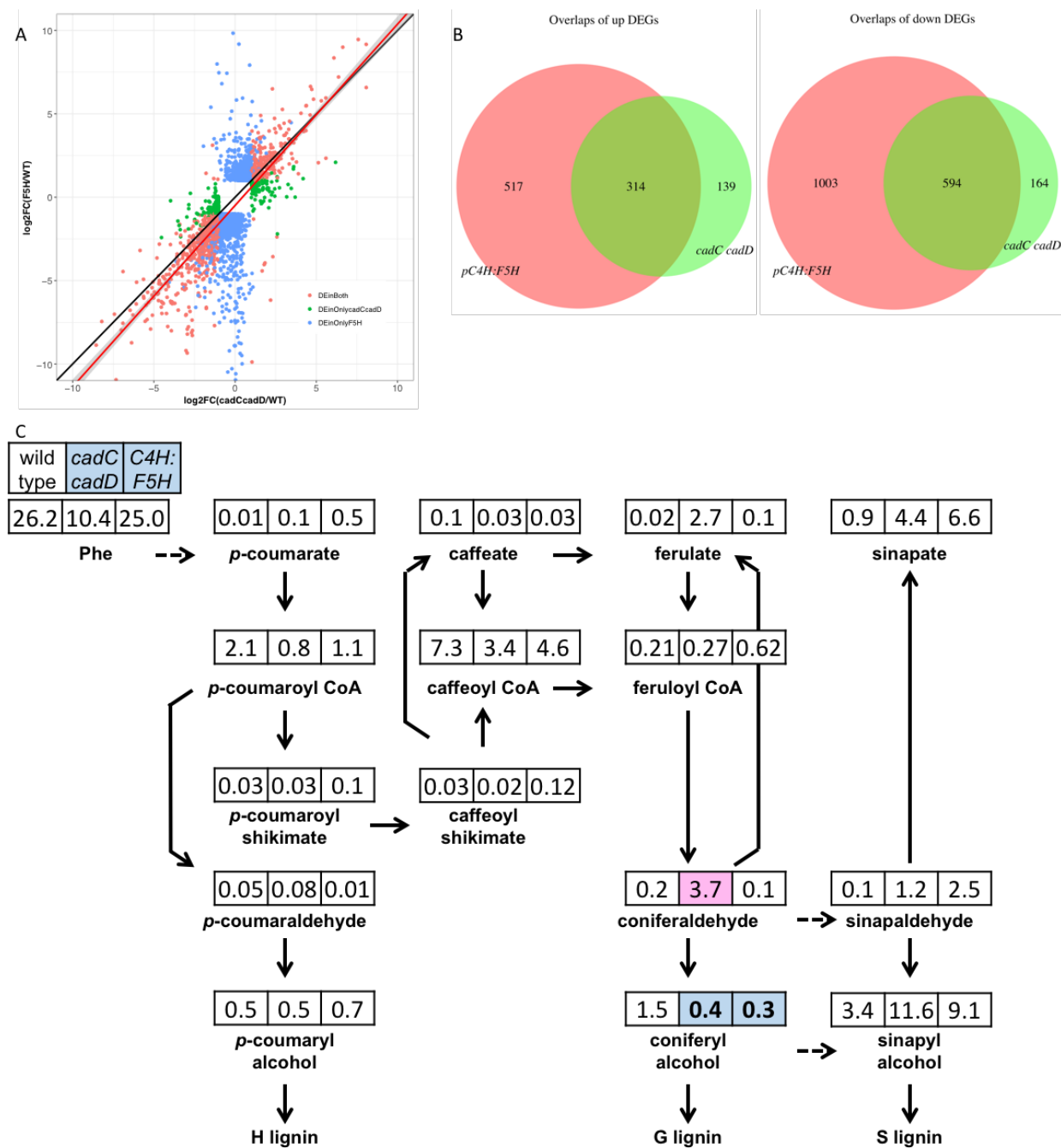


Figure 4.17 Comparison of metabolites and transcripts in wild type, *cadC cadD*, and *C4H:F5H*.

(A) \log_2FC of substantial DEGs in *cadC cadD*, and *C4H:F5H*. (B) Venn diagram of up or down regulated genes in each mutant compared to wild type. (C) Concentrations (in the unit of nmol g FW⁻¹) of phenylpropanoids in wild type, *cadC cadD*, and *C4H:F5H*. Numbers represent means of absolute concentrations of metabolites. One-way ANOVA with Dunnett's test was applied to compare metabolite concentration in each mutant to wild type. Background color indicates the value is significantly higher (pink) or lower (blue) compared to wild type at adjusted *p*-value of 0.05. Bold indicates the value is significant at adjusted *p*-value of 0.01.

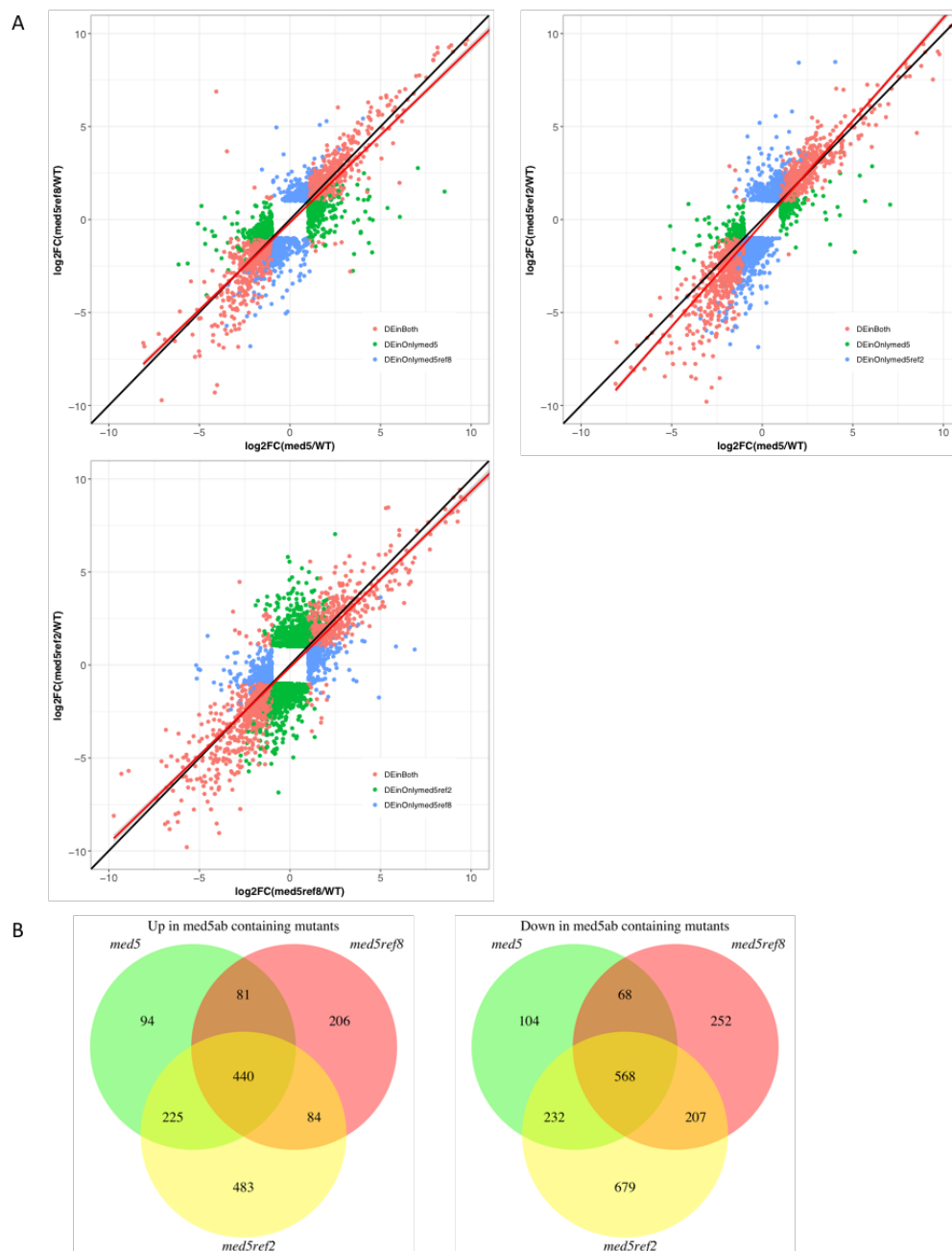
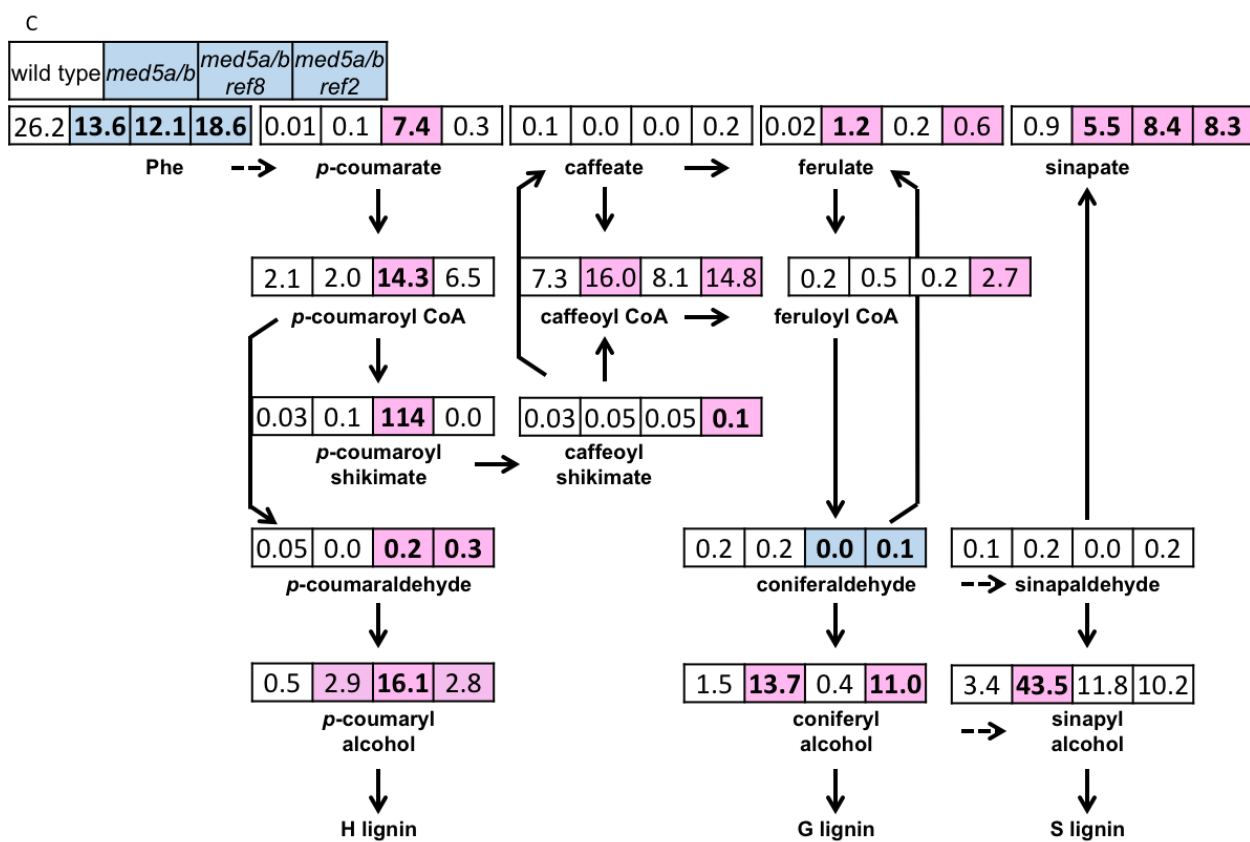


Figure 4.18 Comparison of transcripts and metabolites in wild type, *med5a/b*, *med5a/b ref8*, and *med5a/b ref2*.

(A) \log_2FC of substantial DEGs in *med5a/b*, *med5a/b ref8*, and *med5a/b ref2*. (B) Venn diagram of up or down regulated genes in each mutant compared to wild type. (C) Concentrations (nmol g FW⁻¹) of phenylpropanoids in wild type, *med5a/b*, *med5a/b ref8*, and *med5a/b ref2*. Numbers represent means of absolute concentrations. One-way ANOVA with Dunnett's test was applied to compare metabolite concentration in each mutant to wild type. Background color indicates the value is significantly higher (pink) or lower (blue) compared to wild type at adjusted *p*-value of 0.05. Bold indicates the value is significant at adjusted *p*-value of 0.01.

Figure 4.18 Continued



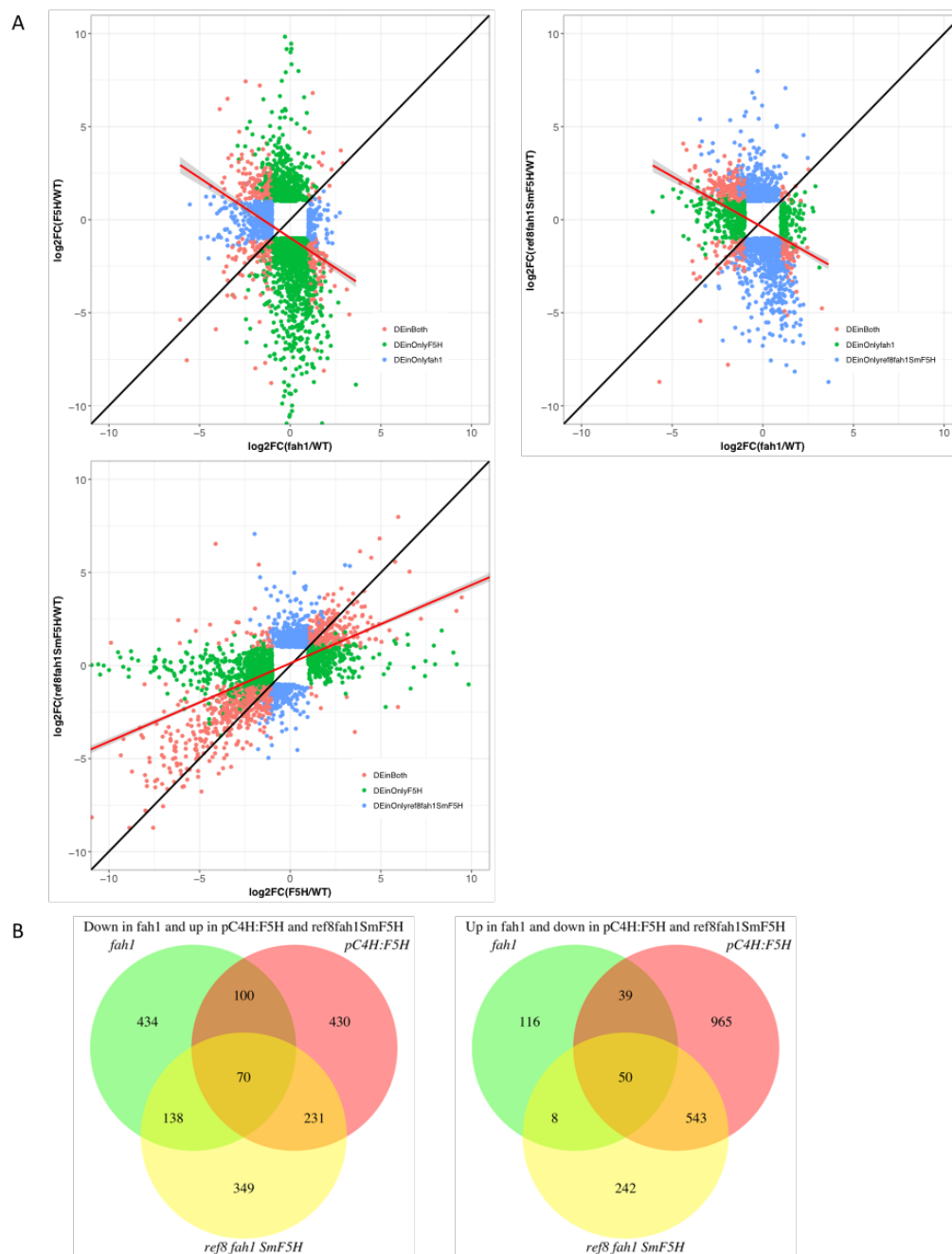


Figure 4.19 Comparison of metabolites and transcripts in wild type, *fah1*, *C4H:F5H*, and *ref8 fah1 SmF5H*.

(A) \log_2FC of substantial DEGs in *fah1*, *C4H:F5H*, and *ref8 fah1 SmF5H*. (B) Venn diagram of up or down regulated genes in each mutant compared to wild type. (C) Concentrations (nmol g FW^{-1}) of phenylpropanoids in wild type, *fah1*, *C4H:F5H*, and *ref8 fah1 SmF5H*. Numbers represent means of absolute concentrations. One-way ANOVA with Dunnett's test was applied to compare metabolite concentration in each mutant to wild type. Background color indicates the value is significantly higher (pink) or lower (blue) compared to wild type at adjusted p -value of 0.05. Bold indicates the value is significant at adjusted p -value of 0.01.

Figure 4.19 Continued

